Molecular genetic analysis of extracellular enzyme secretion by *Erwinia carotovora*

by Philip J. Reeves, B.Sc. (Hons)

A thesis presented for the degree of

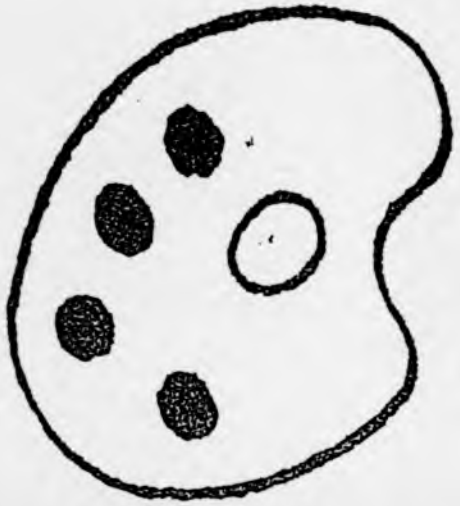
Doctor of Philosophy

Department of Biological Sciences

University of Warwick

September 1991

NUMEROUS ORIGINALS
IN COLOUR



To

my Mother and Father

and

Marian

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgements

I thank George Salmond for his supervision, thoughtful comments and encouragement during the course of this project. Thanks also to Jay Hinton for introducing me to erwinia, and to both Jay and Debs for encouraging my personal growth.

Thanks are due to Dave (Yee-ha) Whitcombe for his intellectual discussions and radical approaches to experimental techniques, Derek Hood for help with DNA methods, Fiona Ellard for trying against all odds to get me organised, Andy Morby and Malcolm M^{C}Crae for much needed help with computers, Stan Tabor for bacterial strains and plasmids, and Dave Hodgson for being so approachable and for answering most of the questions I put to him.

Special thanks to all those from Biological Sciences who contributed to my social life at Warwick, especially Nick, Chris and Ian (sorry about the vodka promotion night).

Most of all, I thank Maz for her constant love, encouragement and support, especially during the preparation of this thesis, and for translating this thesis into English.

I acknowledge the financial support of the SERC.

Declaration

This thesis has been composed by myself and has not been used in any previous application for a degree. The results were obtained by myself, under the supervision of Dr. G.P.C. Salmond, with the exception of those instances where the contribution of others has been acknowledged. All sources of information have been specifically acknowledged by means of reference.

SUMMARY

Erwinia carotovora subsp. carotovora (Ecc) secretes a variety of extracellular enzymes, namely pectinases (Pel), cellulases (Cel) and proteases (Prt). Some of these extracellular enzymes are considered to be the major pathogenicity determinants of this bacterium. Using the chemical mutagen ethyl methyl sulphonate (EMS), a range of Ecc mutants defective in extracellular enzyme production have been generated. One class was found to be pleiotropically defective in the production of Pel and Cel but unaffected for Prt production. Pel and Cel were still synthesised in this class of mutant but both enzymes accumulated within the periplasm. Mutants of the Pel-, Cel-, Prt+ class have been termed Out- mutants. A single Out- mutant, RJP190, was partially resistant to infection by two Ecc bacteriophages. Using a cosmid library of wild-type Ecc, 12 of the 14 Out- mutants were complemented to Out+. Further analysis of the complementing cosmids led to the identification of at least six out loci. A 3.7 kb region of DNA containing out genes was sequenced. This fragment of DNA overlapped with other out genes sequenced in this laboratory. The contiguous DNA (5.7 kb) encoded four proteins, OutD, OutE, OutF and OutG, which were visualised using a T7 directed expression system. The predicted Out proteins were found to share homology with other eubacterial proteins involved in macromolecular trafficking. Accumulated findings strongly suggest that this Out-type system is the major pathway used by Gram-negative bacteria for secreting proteins to the extracellular milieu.

Abbreviations

A	absorbance
AP	ammonium persulphate
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
Cel	cellulase
cfu	colony forming unit
CIP	calf intestinal phosphatase
CMC	carboxymethylcellulose
conc	concentrated
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
ddATP	2',3'-dideoxyadenosine 5'-triphosphate
ddCTP	2',3'-dideoxycytidine 5'-triphosphate
ddGTP	2',3'-dideoxyguanosine 5'-triphosphate
ddTTP	2',3'-dideoxythymidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DMF	dimethylformamide
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTPs	2'-deoxynucleotide 5'-triphosphates
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
Eca	*Erwinia carotovora* subsp. *carotovora*
Ecc	*Erwinia carotovora* subsp. *carotovora*
Ech	*Erwinia chrysanthemi*
E. coli	*Escherichia coli*
EDTA	diaminoethanetetra-acetic acid
EMS	ethylmethylsulphonate
eop	eficiency of plating
ER	endoplasmic reticulum
EtBr	ethidium bromide
EtOH	ethanol

g	gramme(s)
h	hour(s)
IEF	isoelectric focusing
IM	inner-membrane
IPA	isopropanol
k	1000
kb	kilobase(s)
kD	kilo Dalton(s)
l	litre(s)
LB	Luria-Bertani medium
LPS	lipopolysaccharide
M	molar
min	minute(s)
mg	milligramme(s)
ml	millilitre(s)
mM	millimolar
MM	minimal medium
MMA	minimal medium agar
mRNA	messenger RNA
M_w	molecular weight
NA	nutrient agar
NTP	nucleotide 5'-triphosphate
OM	outer-membrane
ONPG	O-nitro-phenol-β-D-galactose pyranoside
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
Peh	polygalacturonase
Pel	pectate lyase
Pem	pectin methylesterase
pfu	plaque forming unit
PGA	polygalacturonate
Pnl	pectin lyase
Prt	protease
psi	pounds per square inch
RNA	ribonucleic acid
RNase	ribonuclease

rpm	revolutions per minute
RT	room temperature
SDS	sodium dodecyl sulphate
sec	second(s)
sp	single species
spp	two or more species
subsp	subspecies
TAE	Tris-acetate-EDTA buffer
TBE	Tris-borate-EDTA buffer
TE	Tris-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TMG	phage buffer
Tris	2-amino-2(hydroxymethyl)-1,3-propane diol
ts	temperature sensitive
u	unit
µg	microgramme(s)
µl	microlitre(s)
µM	micromolar
UV	ultraviolet
V	voltage
v/v	volume/volume
w/v	weight/volume
YE	yeast extract

Amino Acid Abbreviations

A	ALA	Alanine
C	CYS	Cysteine
D	ASP	Aspartic acid
E	GLU	Glutamic acid
F	PHE	Phenylalanine
G	GLY	Glycine
H	HIS	Histidine
I	ILE	Isoleucine
K	LYS	Lysine
L	LEU	Leucine

M	MET	Methionine
N	ASN	Asparagine
P	PRO	Proline
Q	GLN	Glutamine
R	ARG	Arginine
S	SER	Serine
T	THR	Threonine
V	VAL	Valine
W	TRP	Tryptophan
Y	TYR	Tyrosine

Restriction Endonuclease Abbreviations

A	AvaI
B	BamHI
Bg	BglII
C	ClaI
E	EcoRI
H	HindIII
R	EcoRV
S	SalI
Sp	SphI

CHAPTER 1

INTRODUCTION

1.1. Preface

All living cells are organised structures exhibiting high levels of complexity. In order to maintain this state and to ensure the correct functioning of the cell it is imperative that proteins are directed to the correct cellular location. A fundamental question is: how is this targeting achieved? This question can be subdivided into a series of sub-questions. How do cells target certain proteins to specific cellular or extracellular destinations whilst retaining others in the cytoplasm? Do sorted proteins themselves carry sufficient information *per se* or is there a cellular sorting mechanism which selectively segregates them and ensures they reach their appropriate destination? If there is a selective targeting apparatus how does it recognise only certain groups of proteins and what is the nature of this molecular interaction? It is also important to understand how hydrophilic proteins manage to traverse hydrophobic phospholipid bi-layers, especially when they must do this more than once in order to reach their final destination. By answering such questions it should be possible to piece together the overall story and hopefully gain an insight into these fascinating processes.

Initial studies in this area of research concentrated on the targeted protein itself. However, the main aim of this thesis will be to investigate and discuss a new class of proteins which are accessory factors in the cellular targeting of other proteins. There is now considerable evidence to prove the existence of accessory components involved in the targeting of proteins and these will be introduced and discussed in subsequent sections. This work will focus on a group of such proteins involved in the secretion of extracellular enzymes from the Gram-negative bacterium *Erwinia carotovora*

subsp. carotovora. However, in order to set the scene for this story a brief overview of the major breakthroughs in protein targeting is necessary.

1.2. Protein targeting - a global phenomenon

As mentioned earlier the phenomenon of protein targeting is an essential cellular process, and is common to both eukaryotes and prokaryotes. Biological cells, as their name suggests, are essentially self-contained, enclosed, living systems. The presence of a selectively permeable cell envelope allows the containment of the products of metabolic reactions. It is just as essential, however, to ensure the entry of substrates into, and the flow of waste products out of, the cell. Small solutes are able to pass through the semi-permeable membrane. However, the passage of larger molecules is more of a problem. The continued existence of a free living simple cell, such as a bacterium, is dependent upon specialised transport systems allowing the selective passage of certain molecules into and out of the cell. Bacteria are also known to have developed specialised systems enabling the sequestration of scarce ions e.g. iron from the surrounding environment. In order to achieve these functions the correct targeting of proteins to, and beyond, the cell boundary is necessary.

The eukaryotic cell is a complex structure comprising multiple compartments. During evolution, genes encoding organelle proteins are believed to have been relocated from the organelle to the nucleus (Nagley and Devenish, 1989). It is, therefore, necessary that the cognate proteins are re-directed back to their correct locations. Proteins, synthesised in the cytoplasm, are routed to various compartments, which include the endoplasmic reticulum, the Golgi apparatus, organelles, peroxisomes, and the

nucleus. Interestingly, 90% of mitochondrial proteins are encoded allotropically in the nucleus and must therefore be imported (Pfanner et al., 1988). Proteins imported into the mitochondrion, are in some cases, further directed to specific locations within this organelle. Two such proteins, cytochromes c_1 and b_2, which are synthesised as precursors in the cytoplasm, are first of all directed through a fusion point in the two mitochondrial membranes into the matrix (Hartl et al., 1987). They are then transported back across the inner-membrane and into the inter-membrane space. Both of these translocation steps involve proteolytic processing events. The second translocation step might even be similar to the protein export step in bacteria, a process which will be discussed in a later section.

1.3. Protein targeting in eukaryotic systems

Most of the early biochemical work on protein targeting was carried out with eukaryotic systems. A major drawback of using the eukaryotic system, with the exception of yeast, was the lack of a genetic system. This problem was overcome by investigating simpler eukaryotes such as yeast and the genetically amenable prokaryotes. There were, however, advantages in studying protein targeting in higher eukaryotes. Eukaryotic cells are highly specialised structures with distinct organelles performing specific functions. The specialised cells and structures of the eukaryotes simplified the task of isolating components involved in protein targeting.

The aim of this section will be to present a summary of the major discoveries and to introduce these concepts in the form of a brief historical resume.

1.3.1. The 'Palade pathway'

Pioneers in the field of eukaryotic protein targeting were George Palade and co-workers who carried out studies on the secretion of chymotrypsinogen from pancreatic cells of the guinea pig. Some of this early work is described in a review article by Palade (1975) and will be summarised here. Initially, Palade and co-workers used electron microscopy to study the pancreatic cells of the guinea pig. From histological studies in the 1950's, hypotheses were formed about the possible functioning of the pancreatic cells. They were observed to be 'packed with stacked endoplasmic reticulum (ER) cisternae which were studded with ribosomes'. In 1959, using cell fractionation techniques, Palade *et al.* then went on to isolate these cellular components. However, the cell fractionation procedure suffered from limitations and resulted in imperfect separation of components. A major breakthrough was made in 1967 when a combined approach of radiography and an *in vitro* subcellular system was used to follow the secretion of chymotrypsinogen. This method, along with a refined subcellular fraction technique designed by James Jamieson, resulted in the discovery of the 'Palade pathway'. The route of pulse-labelled enzyme was traced from its site of synthesis on ER-attached polysomes, across the membrane and into the lumen of the ER. Chymotrypsinogen was then followed through the ER to the Golgi apparatus and into secretory vesicles, before the final step of expulsion.

The Palade pathway was later verified by an independent piece of work by Novick *et al.* (1981) using the genetically amenable yeast, *Sacharomyces cerevisiae*. A large number of conditionally lethal mutants were generated which were defective in the ability to secrete glycoproteins such

as invertase. The mutants were physiologically analysed at the non-permissive temperature for growth and found to be blocked at various stages of the secretion pathway. Furthermore, all but one of the secretion mutants accumulated exaggerated secretory organelles. By constructing double secretion mutants it became possible to determine the order of secretion events. The order was similar to that previously demonstrated by Palade. In the same piece of work the following features of this process were demonstrated. First, secretory proteins were glycosylated as they entered the ER. Second, at least nine *sec* gene products were required to transfer material to the Golgi apparatus. This process required energy and resulted in further glycosylation of the secretory protein. Third, at least two more functions were required to package the almost fully glycosylated proteins into secretory vesicles. Finally, the transportation of this bud and its fusion to the plasma membrane required energy and a further ten gene products.

After these important discoveries attention was turned to the molecular events of protein targeting. It was clearly demonstrated that secreted proteins were initially synthesised on ER-bound polysomes suggesting that elongation and translocation might be coupled. It was not until 1972, however, that a working model was presented to explain the molecular mechanism by which secreted proteins were directed across the membrane and into the lumen of the ER.

1.3.2. The signal hypothesis

Milstein *et al.* (1972) discovered a slight discrepancy between the size of IgG1 produced *in vitro* when compared to the same immunoglobin produced and secreted *in vivo*. The protein synthesized *in vitro* had a higher molecular

mass and exhibited an altered N-terminal amino acid sequence. From these observations it was speculated that the alteration in the N-terminal region might somehow be involved in the targeting of this protein. It was postulated that this 'leader' sequence might be acting as a signal responsible for protein targeting.

Blobel and Dobberstein (1975) made an important breakthrough by developing an *in vitro* assay for protein secretion. By adding microsomal vesicles to an *in vitro* translation system they managed to couple protein synthesis with its transfer across a membrane. It was demonstrated that proteins were transferred into these microsomal vesicles and were subsequently resistant to proteolytic degradation upon the addition of proteases. Proteins synthesised in the absence of vesicles were slightly larger than those transported into microsomal vesicles. Also, if vesicles were added after the completion of protein elongation, they were not translocated into the vesicles. These results, as well as agreeing with those of Milstein *et al.* (1972), also demonstrated a tight coupling of protein elongation and translocation in this system. It was also proved that the processing observed by Milstein *et al.* (1972) was linked with trans-membrane transport.

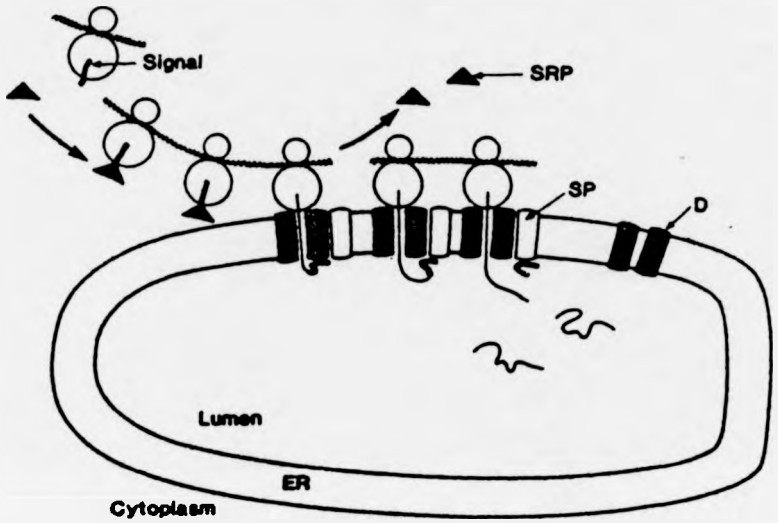
The signal hypothesis suggested that the information for protein translocation across membranes resided in a short 'leader' N-terminal sequence which was removed after transport was complete. The result was a shortened mature form of the protein with an altered N-terminal region.

Walter and Blobel (1980) then went on to isolate the components of the secretion apparatus. By salt washing rough ER membranes, a 250 kD complex was isolated which consisted of six polypeptides. This complex was identified as the signal recognition particle (SRP). The role of this complex

was to bind to the signal-sequence of nascent polypeptides thus preventing further elongation (Walter and Blobel, 1981). Another protein was also discovered which was also released from washed/protease-treated microsomes (Meyer and Dobberstein, 1980). This 72 kD protein restored the elongation of the translationally arrested protein and was called the docking protein (Meyer et al., 1982).

The early events of protein secretion across the ER membrane and into the ER lumen are illustrated in Figure 1.1. This elaborate mechanism is thought to exist to ensure that the secretory protein is translocated across the ER membrane co-translationally, thus preventing it from folding into a translocation-incompetent state. The events leading to protein export in Gram-negative bacteria, and protein secretion in Gram-positive bacteria (which will be discussed in later sections), appear to be similar to the initial steps of secretion in eukaryotes. In both types of prokaryote, translocated proteins are generally synthesised as precursors with N-terminal signal-sequences. Also, some Escherichia coli proteins are known to share sequence similarities to components of the eukaryotic SRP complex (Bassford et al., 1991). However, these proteins (FfH and FtsY) are not thought to be involved in the Sec mediated export pathway in E. coli. The Sec mediated export pathway (also known as the general export pathway [GEP]) is discussed in section 1.4.2. As yet, there is no evidence to suggest that a bacterial SRP exists. Although appearing similar, the mechanisms of export in Gram-negative bacteria (secretion in Gram-positive bacteria) and the initial steps of secretion in eukaryotes, have some differences. The main difference might be the way that proteins are maintained in a translocation-competent state prior to translocation. The way in which prokaryotes achieve this will

Figure 1.1. Protein transport across the endoplasmic reticulum



The signal recognition particle (SRP) interacts with the N-terminal signal-sequence emerging from the ribosome. The ribosome-nascent polypeptide complex is directed to the endoplasmic reticulum (ER) by the SRP and then connects with the membrane-bound docking protein (D). The SRP is then released and protein elongation is restored. The protein is transported through the ER membrane co-translationally and finally cleaved by the signal peptidase (SP). The free polypeptide is released into the lumen of the ER.

From Holland et al. (1986).

be discussed in sections 1.4.2.3. and 1.4.2.4.

1.4. Protein targeting in the prokaryotes

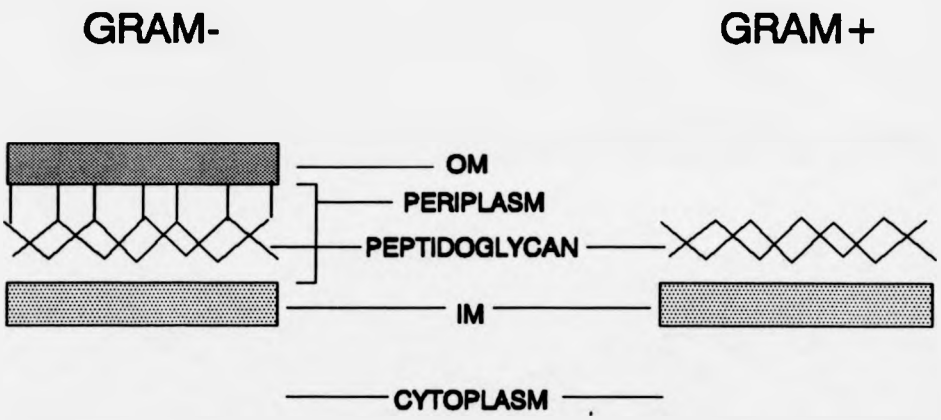
1.4.1. Introduction

Bacterial cells, although less complex than eukaryotic cells, must direct proteins to specific areas within and beyond the cell. The cellular contents of a Gram-positive bacterium are enclosed by a cell envelope comprising a single lipid bi-layer, peptidoglycan and a surface protein array (Pugsley and Schwartz, 1985). The cell envelope of a Gram-negative bacterium is more complicated due to an additional membrane. Diagrams showing the main features of the two types of bacterial cell envelope are illustrated in Figure 1.2. A more detailed diagram of the Gram-negative bacterial cell wall is given in Figure 1.3.

Gram-positive and Gram-negative bacteria secrete proteins into the extracellular environment. They also direct proteins to their cell envelope. Proteins secreted from Gram-positive bacteria must traverse a single membrane. Protein secretion by Gram-negative bacteria is more complicated because two distinct membranes must be traversed. This fact has created much confusion in the literature with the use of the terms 'protein export' and 'protein secretion'.

In this thesis the term 'protein secretion' is used to describe the transfer of a protein to an extracellular location. 'Protein export' will be used to describe the passage of a protein across the IM into the periplasm of a Gram-negative bacterium. Proteins which are processed during translocation (by the removal of a signal-sequence) will be referred to as 'pre-proteins' in

Figure 1.2.
Simplified diagramatic representation of the Gram- and Gram+ bacterial cell walls

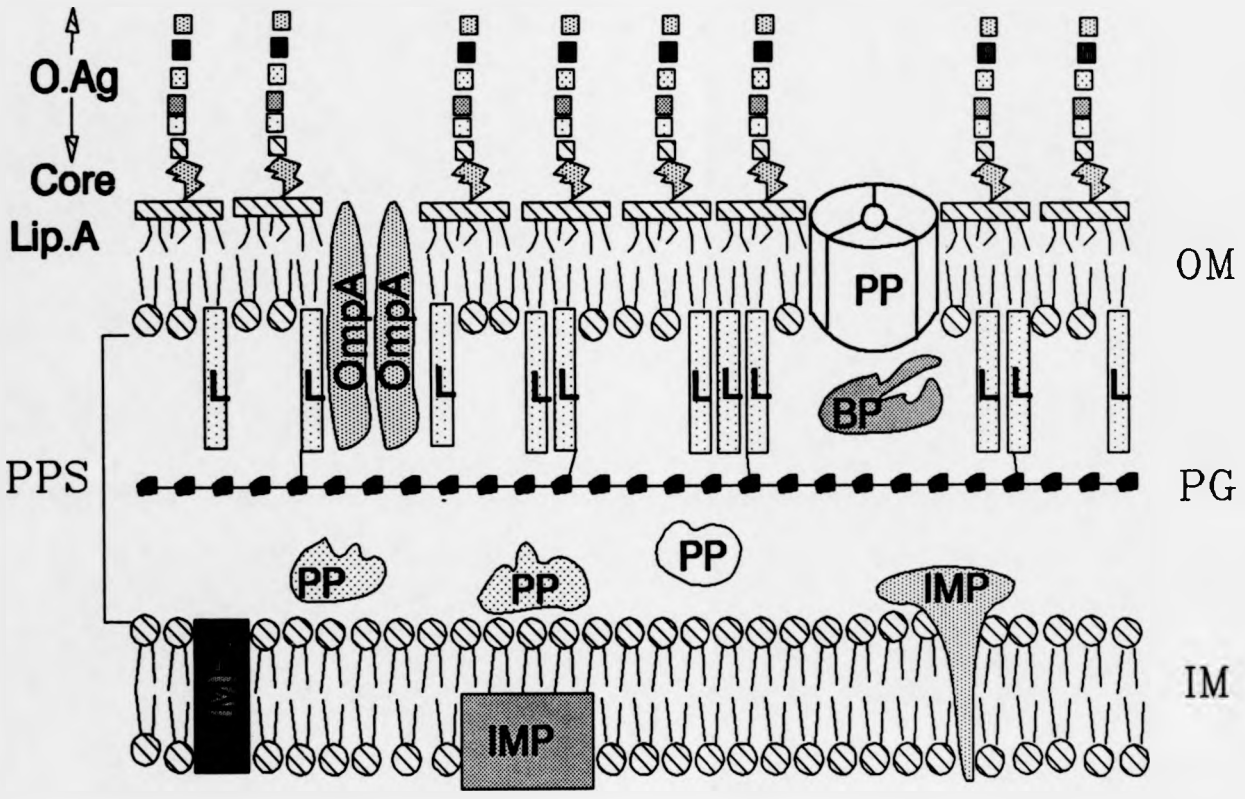


From Pugsley (1988).

Legend for Figure 1.3.

Structure of the E. coli cell envelope from Lugtenberg and van Alphen (1983). BP = periplasmic binding protein; IM = inner-membrane; IMP = inner-membrane protein; L = lipoprotein; LipA = lipid A; LPS = lipopolysaccharide; O.Ag = O-antigen; OM = outer-membrane; PG = peptidoglycan; PP(PPS) = periplasmic protein; PPS = periplasm; PP(OM) = pore forming protein trimer (porin). The periplasm also contains membrane-derived anionic oligosaccharides (Hancock, 1991). Although the peptidoglycan is drawn as a single-layer, it is now thought to be cross-linked to form a gel.

Figure 1.3 The Gram-negative cell wall



order to distinguish them from the mature forms of that particular protein.

Protein secretion in Gram-positive and Gram-negative bacteria will be discussed in detail in sections 1.6. and 1.7. First, the process of protein export in E. coli will be presented. Protein 'export' describes the transfer of proteins across the IM into the periplasm of this Gram-negative bacterium. This process has been investigated intensely and much is known about the components of the E. coli export machinery. The equivalent process (protein translocation across a single lipid bi-layer) is sufficient for protein secretion in Gram-positive bacteria. Indeed, β-lactamase from E. coli is secreted from Bacillus subtilis (Kallio et al., 1986). Also, as will be discussed in section 1.7.6., one of the mechanisms by which Gram-negative bacteria secrete proteins is thought to include export followed by a second step resulting in complete secretion.

1.4.2. Protein export in E. coli

1.4.2.1. Introduction

This area of work has been intensely studied and has been the subject of some recent reviews (Randall et al., 1987; Saier et al., 1989; Bassford et al., 1991; Bieker and Silhavy, 1990). The components of the E. coli export apparatus and their possible functions will be described. Some elaborate genetic techniques which were used in this investigation will also be described.

1.4.2.2. Components of the E. coli export machinery

The E. coli export pathway has been analysed using a combination of

genetic, biochemical and physiological techniques. Table 1.1. lists the known components of the E. coli export apparatus and their putative modes of action.

A genetic approach was used to isolate mutants defective in protein export. The approaches taken have recently been reviewed (Bieker and Silhavy, 1990) and are described below. Hybrid proteins were constructed between the 3' region of the lacZ gene of E. coli and the 5' end (containing the signal-sequence encoding DNA) of genes encoding for the normally exported proteins malE and lamB. The hybrid proteins (MalE-LacZ and LamB-LacZ) were targeted to the IM of E. coli and their production was induced by maltose. The fusion proteins were recognised by the export apparatus but could not complete the process because the β-galactosidase molecule adopted a conformation that could not be exported. This led to an inactive β-galactosidase (because it was unable to form functional tetramers) and also resulted in the jamming of the export apparatus. High levels of production of either of these hybrid proteins (controlled by maltose levels) were lethal as they completely jammed the export apparatus. When these fusions were introduced into a lac Δ strain of E. coli, the resulting phenotype was Lac- and Mal^S. Maltose resistant (Mal^r) mutants were then searched for in order to identify intragenic targeting signals for the fusion proteins. Such mutants would not target the hybrid protein to the IM and therefore would not jam the export machinery. Indeed, the vast majority of the mutations (conferring Mal^r) were located in the signal-sequence of the hybrid proteins.

Three approaches were then taken to identify components of the export apparatus. The first approach was to identify extragenic suppressors of

Table 1.1. Components of the *E. coli* protein export apparatus

Protein	Location	Size (kD)	Function
SecB	Cytoplasm	12	Form complexes with pre-proteins to maintain translocation competence [Chaperones]
Trigger factor		60	
GroEL		910	
SecA	Membrane (peripheral)	102	ATPase, directs [pre-protein/chaperone] complex to IM
SecY	Membrane (integral)	49	Interact with SecA, translocators?
SecE		14	
Lep	Membrane (integral)	36	Process pre-proteins
LspA		18	Process pre-lipoproteins
SecD	Membrane (integral)	65	Unknown, late step in translocation?
SecF		35	

Legend

The information for this table was obtained from the following sources; Bieker and Silhavy, (1990); Bassford *et al.*, (1991), Lecker *et al.*, (1989) and Crooke *et al.*, (1988).

signal-sequence mutations. Strains carrying mutations in the signal-sequences of either MalE or LamB were used. Restoration of the Mal+ phenotype was sometimes produced by mutations in export proteins which interacted with the (mutated) signal-sequence of MalE or LamB. These genes were termed prl as they were involved in protein localisation. Three genes were identified which were called prlA, prlD and prlG. Surprisingly, the prl mutations were not lethal even though they were in essential genes.

The second approach taken led to the identification of the sec genes, some of which were allelic with prl genes. In this approach general export defects were screened for in the fusion protein (MalE-LacZ) in a lacΔ background (as described earlier). Secretion mutants of the above strain were identified on the basis of increased β-galactosidase activity at 30°C. The production of an active β-galactosidase was in some cases expected to result from mutations in the export machinery. Such mutations would cause an inability of the export apparatus to recognise the signal-sequence of the MalE-LacZ protein which would then reside (and form active molecules) within the cytoplasm. Such mutants were then re-screened for a conditionally lethal phenotype. This approach was used to identify secA and secB. A similar approach, using PhoA-LacZ and LamB-LacZ fusions, was used to identify secD and secE.

The third approach came from findings that secA gene expression was de-repressed under conditions that inhibited protein export. A SecA-LacZ fusion was constructed (in a merodiploid containing secA+) and conditional lethal mutants were isolated with raised levels of β-galactosidase activity. This approach led to the discovery of secE (cold-sensitive). Mutations were also found in the fusion protein as well as, secA, secY, secD and secF. No

mutations in secB were isolated because secB mutations do not cause over-expression of secA.

Biochemical studies have focussed on isolating the components of the export apparatus. The reconstitution of an in vitro synthesis/transport system then followed which enabled the study of the individual components of the export apparatus. The in vitro system has relied heavily on the use of inverted plasma membrane vesicles. Under the correct conditions and with the necessary components it is possible to direct various, normally exported, proteins into such vesicles (Swidersky et al., 1990). Once proteins are internalised into vesicles they are immune from proteinase attack. This feature can serve as an assay to monitor the progress of protein translocation. This type of experiment has been used to demonstrate the necessity of SecA (Swidersky et al., 1990, Cunningham et al., 1989), SecB (Watanabe and Blobel, 1989) and SecY/E (Brundage et al., 1990) in protein export. A summary of the components of the E. coli export machinery and their proposed functions is shown in Table 1.1. and described below.

The translocation of proteins across the IM of E. coli involves several cytoplasmic and membrane protein factors (see above). These include six sec gene products (SecA [PrlD], SecB, SecD, SecE [PrlG], SecF, SecY [PrlA]) and signal-peptidase(s). Other proteins, as well as the Sec proteins, have been implicated in the export of some classes of protein. Trigger factor has been demonstrated to be needed for the in vitro 'export' of pre-OmpA into E. coli IM vesicles (Crooke and Wickner, 1987). The E. coli heat shock proteins, GroEL and GroES, have also been implicated in protein export in E. coli (Kusukawa et al., 1989).

Biochemical approaches have been used to analyse the functions of

SecA and SecB and have been reviewed (Bieker and Silhavy, 1990). An elaborate genetic approach, termed suppressor-directed inactivation (SDI), was used to investigate SecY and SecE. The LamB-LacZ hybrid protein with a defective signal was used in a merodiploid E. coli strain carrying a suppressing SecE (recognising the mutated LamB-LacZ signal-sequence) protein and SecE+. The cell functions normally because the SecE+ protein does not recognise (and is not blocked by) the mutant hybrid protein. The hybrid protein is, however, trapped by the mutated (suppressing) SecE. The suppressor directed inactivation step can be studied to identify the stage of the block. The signal-sequence of The LamB-LacZ protein was not processed when blocked at the stage of action of SecE. However, when blocked at SecY, the protein had been processed. This indicated that SecY functioned at a later stage in the export pathway than SecE. Recent work has suggested that a truncated version of SecE is functional (Schatz et al., 1991). The SecE truncate had only one of three membrane-spanning domains remaining and this was sufficient for its function.

In addition to these proteinaceous factors, ATP (Lill et al., 1989) and the proton motive force (Schiebel et al., 1991) are also required for protein export across the E. coli inner-membrane (IM).

1.4.2.3. Exported proteins must be translocation competent

Proteins, which are synthesised as single chain amino acid polymers, must at some stage fold in order to attain their 'functional' conformation. It has been suggested that newly synthesised proteins in the environment of the 'aqueous' cytoplasm will fold in order to internalise hydrophobic residues and produce a more thermodynamically favourable conformation (Wickner,

1989). This poses a problem for a protein which is to be transported across a membrane because it is important that it does not fold in such a way that will prevent its translocation. Some theories on possible mechanisms of protein translocation across membranes will be discussed in the next section. However, it is generally assumed that proteins which are transported across the IM are in a relaxed conformation. It has been suggested that this transfer does not proceed instantaneously or continuously, but occurs by segments of the polypeptide chain (Singer et al., 1987). If this mechanism is correct then the protein must be transferred co-translationally or remain in a relaxed (translocation competent) conformation before traversing the membrane. Alternatively, the protein might adopt its native conformation within the cytosol and then unfold prior to translocation. Proteins are unfolded when they are imported into mitochondria (Eilers and Schatz, 1986). Some possible mechanisms of protein translocation will be briefly discussed in section 1.5.

Proteins might remain translocation competent by several mechanisms. One possibility, as mentioned earlier (section 1.3.2.), would be to translocate co-translationally. Another mechanism might involve other proteins which prevent the transported protein folding into a translocation incompetent conformation.

1.4.2.4. The chaperones

A group of proteins which are thought to be involved in 'stabilising' pre-proteins to maintain a translocation competent conformation have been identified. These proteins have been termed 'chaperones' (Ellis, 1987) and in E. coli are thought to include SecB (Watanabe and Blobel, 1989), trigger

factor (Crooke and Wickner, 1987) and GroEL/GroES (Kusukawa *et al.*, 1989; Altman *et al.*, 1991). SRP in the ER of eukaryotes and some of the yeast heat shock proteins (*hsp*70 products) have also been proposed to perform similar roles (Meyer, 1988).

As described in section 1.4.2.3., SecB is thought to interact with some exported proteins. It has been proposed that this protein prevents premature folding of newly synthesised proteins into translocation incompetent conformations (Collier *et al.*, 1988). The nature of the interaction between SecB and the pre-protein is not clear. One pair of workers suggest that SecB binds to the signal-sequence of pre-proteins (Watanabe and Blobel, 1989). Another group, however, proposed that the signal-sequence modulated the folding of the protein so that SecB could bind to other regions of the protein (Randall *et al.*, 1990).

Trigger factor is a protein of M_w 63 kD which maintainins ProOmpA (an *E. coli* OM protein) and possibly other pre-proteins, in a translocation-competent form (Crooke and Wickner, 1987; Crooke *et al.*, 1988).

The heat-shock proteins (GroES and GroEL) are also involved in protein export in *E. coli*. Temperature-sensitive *groEL* and *groES* mutants were used to demonstrate the requirement of these proteins for the export of β-lactamase (Kusukawa *et al.*, 1989). These workers also showed that the export of other proteins including PhoA, MalE and OmpA was only minimally affected in these mutants.

It appears that SecB, trigger factor and the heat shock proteins all play a part in protein export in *E. coli* by maintaining translocation-

competence in pre-proteins. Furthermore, it seems that these different chaperones might recognise certain groups of pre-proteins.

1.5. The mechanics of protein translocation across membranes

Studies on the protein translocation systems of both prokaryotes and eukaryotes have resulted in the isolation of the components involved. It has also been possible to reconstitute both systems *in vitro*. However, even with this information, it is still not known how a hydrophilic protein passes through a lipid bi-layer with an hydrophobic interior. Assuming that the protein is translocated as a linear molecule, it might pass directly through the lipid bi-layer or through a proteinaceous pore. One hypothesis is the helical hairpin hypothesis which favours the transfer of the protein while it is in direct contact with the membrane (Engelman and Steitz, 1981). Engelman and Steitz propose that segments of the polypeptide undergoing export spontaneously transfer by pairing with the hydrophobic α-helix formed by the signal-sequence.

Singer *et al.*, (1987) favour the model in which the protein passes through a pore and argue that it would be thermodynamically unfavourable for polar amino acids to come into direct contact with the hydrophobic interior of the lipid bi-layer.

Studies using M13 (the filamentous coliphage) have helped to elucidate the mechanism of protein translocation across membranes. The M13 phage coat protein is thought to interact directly with the IM, via electrostatic interactions between positively charged regions of the phage coat protein and the negatively charged phospholipid head groups of the IM (Kuhn *et al.*, 1990a and b). A transmembrane loop formed by the phage coat protein leads

to the insertion of the protein into the IM. After membrane insertion, cleavage of the signal-sequence occurs. It was originally thought that the negatively charged residues of the phage coat protein were involved in its (electrophoresis-like) transfer across the membrane towards the positively charged outer face of the membrane. Kuhn et al. (1990) substituted the acidic residues at positions +2, +4 and +5 in the phage protein with positively charged (arginine) residues. Membrane insertion by this altered protein still occurred, albeit at a slower rate. The theory of electrophoresis-like transfer of the phage coat protein through the membrane was therefore disproved.

1.6. Protein targeting in Gram-positive bacteria

1.6.1. Introduction

The secretion of proteins from Gram-positive bacteria is widespread (Table 1.2.). Proteins secreted from Gram-positive bacteria must traverse one lipid bi-layer before reaching the extracellular membrane. They are then released into the extracellular milieu through the matrix of the cell wall. The equivalent process in a Gram-negative bacterium would result in the transfer of the ('exported') protein to the periplasm. A recent study showed that antibodies raised against two *E. coli* export proteins (SecA and SecB) did not cross react with any proteins from a number of different distantly related bacteria including *Bacillus cereus* (de Cock and Tommassen, 1991). This finding suggests that, although there are functional similarities between the export apparatus of *E. coli* and the secretion apparatus of bacilli species, there is no similarity in the components of them (at least for SecA and SecB).

The task of protein secretion for a Gram-positive bacterium is relatively simple compared to a Gram-negative bacterium. This might explain why the Gram-positive bacteria are more prolific producers of extracellular enzymes than Gram-negative bacteria (compare Tables 1.2. and 1.4.). The best characterised of the Gram-positive bacteria, both genetically and physiologically, are the *Bacillus* species. *Bacillus* species are especially important in producing a large number of important enzymes. Some of these are listed in Table 1.3. Although the bacilli produce large quantities of extracellular enzymes, a feature which has been exploited industrially for many years (Arbige and Pitcher, 1989), the mechanism of enzyme secretion is not well understood.

Table 1.2. Proteins produced by some Gram-positive bacteria.

Bacterium	Protein produced
Bacillus subtilis	∝-Amylase Subtilisin Neutral protease β-glucanase Levanosucrase
Bacillus licheniforms	Lipopenicillinase Alkaline phosphatase
Bacillus amyloliquiquefaciens	∝-Amylase Subtilisin Neutral protease Alkaline protease
Bacillus cereus	β-lactamase (type3) β-lactamase (type1)
Bacillus coagulans	Amylase
Staphylococcus aureus	Lipopenicillinase Enterotoxin B ∝-toxin Protein A
Corynebacterium diphtheriae	Diphtheria toxin
Clostridium thermocellum	Endoglucanase A

The information for this table was taken from Pugsley and Schwartz (1985).

Table 1.3 Industrial enzyme production by *Bacillus* species

Enzyme	Source	Application
Bacterial α-amylase	*Bacillus subtilis* *Bacillus licheniformis*	Starch conversion
β-Glucanase	*Bacillus subtilis* *Bacillus licheniformis*	Brewing and food processing
Neutral protease	*Bacillus subtilis*	Brewing/flavouring
Alkaline protease	*Bacillus licheniformis*	Detergents

From Arbige and Pitcher (1989).

Most proteins secreted by Gram-positive bacteria are synthesised as precursors with N-terminal signal-sequences. Signal-sequences of proteins secreted by Gram-positive bacteria are between 28 and 34 amino-acids in length (Borchert and Nagarajan, 1991) and, are therefore, longer than those found in proteins exported/secreted by Gram-negative bacteria, which are typically 20 amino acids in length. Gram-positive signal-sequences have two to five positively charged residues at the extreme N-terminal region, a hydrophobic core (12-18 residues), and a signal peptidase processing site (Borchert and Nagarajan, 1991).

Most proteins secreted from bacilli require multiple processing events (Borchert and Nagarajan, 1991). Two examples are protease and RNase which both contain a pro-peptide between the signal-sequence and the mature protein. The signal-sequence directs the pro-protein to the surface of the cell and the proteolytic processing of the pro-peptide results in the release and activation of the secreted protein. This mechanism might exist to ensure that potentially toxic proteins are not active within the cell prior to their secretion. Proteins that are secreted from Bacillus sp. with a single processing step include levansucrase (Borchert and Nagarajan, 1991) and -amylase (Kontinen et al., 1991).

1.6.2. The genetics of protein secretion in bacilli

Mutants of Gram-positive bacteria which are defective in the ability to secrete extracellular enzyme have been described. The Sec- mutants of Bacillus licheniformis accumulated alkaline phosphatase at specific sites in the cell envelope (Kumar et al., 1983). Bacillus subtilis mutants which are pleiotropically defective in the secretion of several extracellular enzymes have been generated (Kontinen and Sarvas, 1988). These mutants (termed

prs) mapped to four separate loci on the B. subtilis chromosome. The prs mutants were isolated by screening for an α-amylase defective phenotype. Some of the prs mutants were pleiotropically defective in the production of other extracellular enzymes including, protease and lipopenicillinase. Other phenotypes displayed by some classes of prs mutants were: inability to sporulate (although the respective mutations did not map to any previously known spo loci); inability to be transformed (prs-26, prs-13 and prs-33); and decreased motility (prs-13). Three prs mutants (prs-3, prs-29 and prs-40) mapped to a single locus termed prsA (Kontinen et al., 1991). This gene encoded a lipoprotein involved in a late stage of protein secretion in B. subtilis. Furthermore, PrsA was similar in sequence to PrtM, a protein involved in the maturation of a secreted proteinase (which contains a pro-peptide) from Lactococcus lactis. Kontinen et al. (1991) proposed that PrsA and PrtM might be 'extracellular' molecular chaperones which assist the folding of secreted proteins.

1.7. Protein targeting in Gram-negative bacteria

1.7.1. Introduction

The Gram-negative bacteria secrete a range of different proteins, some of which are listed in Table 1.4. The structure of the Gram-negative bacterial cell wall is illustrated in Figure 1.3. It comprises two distinct lipid bi-layers, the inner-membrane (IM) and the outer-membrane (OM), which are separated by a space called the periplasm. The OM is a structurally-unique membrane in bacteria because it is an asymmetric bi-layer (Hancock, 1991). The inner-layer (periplasm proximal) is composed of phospholipids whereas the outer-layer is predominantly composed of lipopolysaccharide (LPS). The

Table 1.4. Proteins produced by some Gram-negative bacteria

Bacterium	Protein produced
Escherichia coli	Heat labile enterotoxin Heat stable enterotoxin ∝-haemolysin Colicins
Citrobacter freundii	Endo-β-D-galactosidase Colicin A
Salmonella typhimurium	Colicins
Klebsiella pneumoniae	Pullulanase
Serratia spp.	Proteases Nuclease (RNA and DNA) Lipase Colicin L
Erwinia spp.	Pectinases Cellulases Proteases
Vibrio cholerae	Cholera toxin ∝-Haemolysin
Vibrio spp.	Protease Collagenase
Pseudomonas aeruginosa	Phospholipase C (haemolysin) Toxin A Alkaline phosphatase Staphylolytic enzyme Protease Elastase Pyocins (S type) Alginase
Pseudomonas spp.	Agarase Protease Poly-β-hydroxybutyrate depolymerase
Haemophilus influenzae	IgA protease
Neisseria gonorrhoeae	IgA protease

Table 1.4. (cont.)

Aeromonas hydrophila	Protease Haemolysin Acyltransferase
Yersinia enterolytica	Heat stable enterotoxin
Bacteroides fragilis	Endo-β-D-galactosidase
Bordetella pertussis	Adenylate cyclase
Myxococcus xanthus[a]	Protease Amylase DNase RNase Bactriolytic proteins

Information was taken from Pugsley and Schwartz (1985) except (a) which is from Nicaud et al. (1984).

LPS is responsible for the barrier properties of the OM. The OM also contains proteins and lipoproteins. The lipoproteins are a predominant class of OM proteins. The other major class contain a high content of β-sheet structures (Hancock, 1991). The periplasm is not a 'space' as the old name 'periplasmic space' suggests, but contains proteins, anionic oligosaccharides and also peptidoglycan. The peptidoglycan layer is not a rigid structure as often thought, but has a gel-like consistency comprising cross-linked peptidoglycan molecules (Hancock, 1991).

The cell envelope must be traversed by proteins secreted by Gram-negative bacteria. This appears to be a formidable obstacle for secreted proteins. However, it is now becoming clear that numerous mechanisms have evolved to enable the selective passage of proteins through this structure.

Two models have been proposed to describe the major routes taken by proteins secreted from Gram-negative bacteria. These models are illustrated in Figure 1.4. and examples are given in the following section.

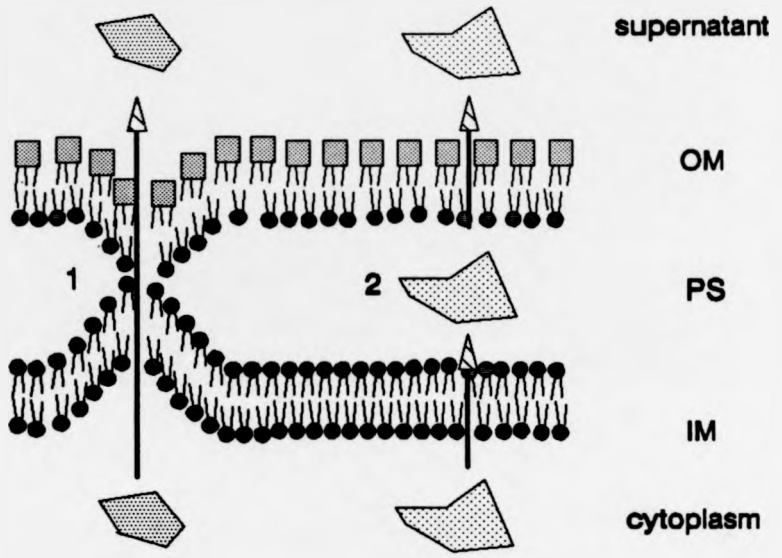
In the first model, the secreted protein passes through a fusion point, encompassing the IM and OM, in a single step. The point of contact (pore?) where the IM and OM fuse is thought to be maintained by proteinaceous factors. These proteins might also be actively involved in the translocation process.

The second model is a two-step process where the protein to be secreted is initially synthesised as a pre-protein with a classical signal-sequence. The first step results in the export of the protein across the IM into the periplasm via the general export pathway (section 1.4.2.). The pre-protein is processed during translocation by the action of a signal peptidase. In the second step the protein, which may remain transiently in the periplasm, is secreted across the OM into the extracellular milieu.

Legend for Figure 1.4.

Figure 1.4 shows the two possible routes which might be taken by proteins secreted by Gram-negative bacteria. The first route (1) shows how a secreted protein might traverse the IM and OM in a single step through a fusion point in the two membranes. Examples of proteins which are secreted by a one-step mechanism are discussed in section 1.7.4. The second secretion route described in this diagram (2) consists of two-steps. The protein to be secreted is first exported across the IM into the periplasm. This transient periplasmic intermediate requires a further translocation step across the OM before it is secreted. Examples of proteins which are thought to be secreted in this manner are discussed in section 1.7.6.

Figure 1.4. Models for protein secretion by Gram-negative bacteria



Alternative models have been proposed to explain the secretion of some proteins from Gram-negative bacteria and will be described in the following section.

1.7.2. Secretion of bacteriocins

Bacteriocins are narrow host range toxins and are widespread among both Gram-positive and Gram-negative bacteria (Harkness and Olschlager, 1991). The best characterised of the bacteriocins are those produced by *E. coli* and are known as colicins. Colicins are single polypeptides encoded as part of a three gene operon which is usually plasmid borne. The other two genes encode an immunity protein (for host cell protection) and a lysis protein (to aid release of cytoplasmically-accumulated colicin). The lysis gene encodes a small protein of 5 kD which is proposed to destabilise the bacterial cell envelope by activating the detergent resistant phospholipase (encoded by *pldA*). PldA is thought to alter the composition of membrane phospholipids resulting in the general release of periplasmic proteins (Harkness and Olschlager, 1991). The release of the colicin to the environment is slow and is accompanied by a partial decrease in the optical density ('quasi lysis') of the culture (Howard *et al.*, 1991).

Recent findings have suggested that for the release of colicin A from *E. coli*, the lysis protein (CalA) causes phospholipase A independent alterations in the integrity of the cell envelope (Howard *et al.*, 1991). The precise nature of the membrane perturbation caused by CalA is not understood.

Alternative models have been proposed to explain the secretion of some proteins from Gram-negative bacteria and will be described in the following section.

1.7.2. Secretion of bacteriocins

Bacteriocins are narrow host range toxins and are widespread among both Gram-positive and Gram-negative bacteria (Harkness and Olschlager, 1991). The best characterised of the bacteriocins are those produced by *E. coli* and are known as colicins. Colicins are single polypeptides encoded as part of a three gene operon which is usually plasmid borne. The other two genes encode an immunity protein (for host cell protection) and a lysis protein (to aid release of cytoplasmically-accumulated colicin). The lysis gene encodes a small protein of 5 kD which is proposed to destabilise the bacterial cell envelope by activating the detergent resistant phospholipase (encoded by *pldA*). PldA is thought to alter the composition of membrane phospholipids resulting in the general release of periplasmic proteins (Harkness and Olschlager, 1991). The release of the colicin to the environment is slow and is accompanied by a partial decrease in the optical density ('quasi lysis') of the culture (Howard *et al.*, 1991).

Recent findings have suggested that for the release of colicin A from *E. coli*, the lysis protein (CalA) causes phospholipase A independent alterations in the integrity of the cell envelope (Howard *et al.*, 1991). The precise nature of the membrane perturbation caused by CalA is not understood.

1.7.3. Release of bacteriophages

The release of bacteriophages from the host cell is usually accompanied by cell lysis resulting from the actions of bacteriophage encoded lysis functions (Pugsley, 1988). For example, bacteriophages MS2 (Walderich and Holtje, 1989) and ØX174 (Witte et al., 1990a) encode lysis proteins which are thought to encourage the formation of a pore between the IM and OM. In both cases the formation of pores have been demonstrated and visualised using immunolabelling (Walderich and Holtje, 1989) and electron microscopy (Witte et al., 1990b). Unlike colicin release, however, cytoplasmic proteins are also released and the cells subsequently lyse.

The release of filamentous E. coli bacteriophages such as M13 does not result in host cell lysis (Pugsley et al., 1990a). Filamentous bacteriophages are continuously assembled and released without killing the host cell (Russel, 1991). One of the bacteriophage encoded proteins (geneIV product) is synthesised as a precursor with a signal-sequence and uses the E. coli export machinery before inserting into the OM (Russel, 1991). The GeneIV protein might form a pore through which the maturing filamentous bacteriophage is extruded (Russel, 1991).

1.7.4. The one-step secretion mechanism

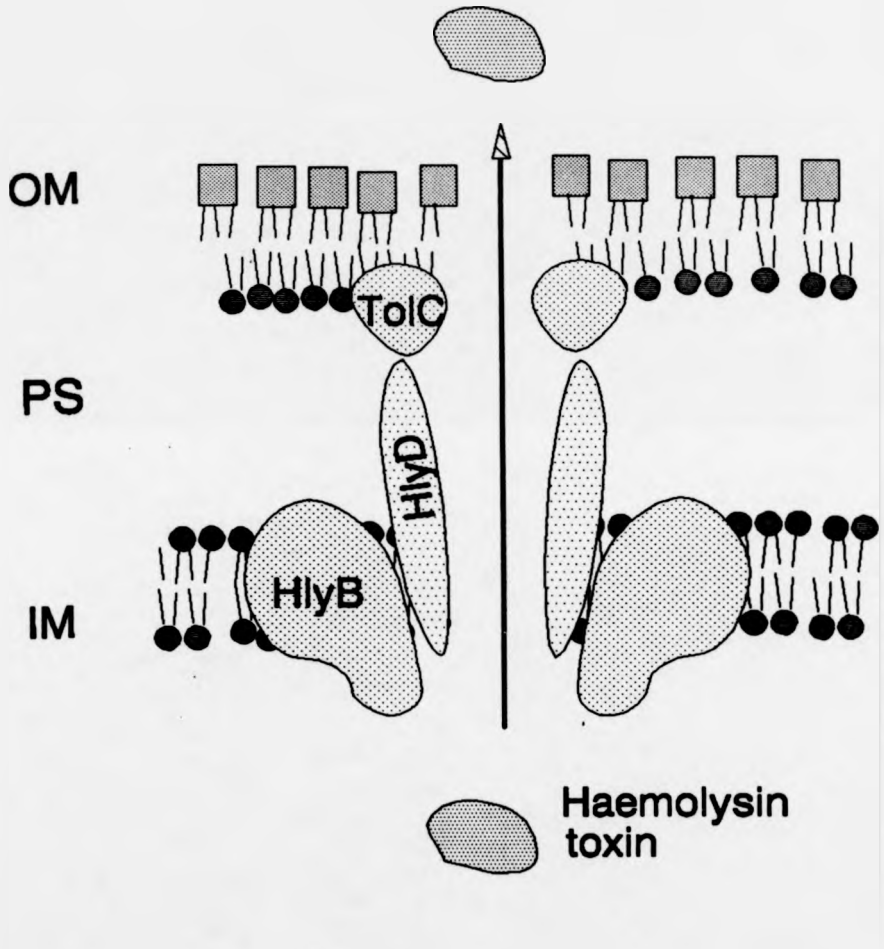
Hemolysin is secreted simultaneously across the IM and OM of E. coli as described by the one-step model in Figure 1.5. (Holland et al., 1990). This event occurs without the processing of an N-terminal signal-sequence and without the presence of a detectable periplasmic intermediate (Holland et al., 1990). The genetic organisation of the hly genes and the proposed arrangement of the Hly proteins in the E. coli cell envelope is shown in Figure 1.5. This model has been updated from Holland et al. (1990) using

Legend for Figure 1.5.

The secretion of hemolysin toxin (HlyA) is dependent on three protein factors, HlyB, HlyD and TolC (Holland et al., 1990; Wandersman and Delepelaire, 1990). HlyB and HlyD co-fractionate with the IM (Holland et al., 1990) whereas TolC is an OM protein (Wandersman and Delepelaire, 1990). HlyB has a potential ATP binding domain which is found in a large family of membrane translocators (Holland et al., 1990). This family includes eukaryotic proteins such as the Mdr family which are produced by drug-resistant tumour cells (Gerlach et al., 1986) and CFTR, a protein that is defective in humans with cystic fibrosis (Riordan et al., 1989).

Secretion of HlyA proceeds without a detectable periplasmic intermediate. The secretion of HlyA and other bacterial Hly-like proteins is discussed in section 1.7.4.

Figure 1.5. Haemolysin secretion



data from Wandersman and Delepelaire (1990). The findings by Wandersman and Delepelaire (1990) are described below.

The E. coli hemolysin structural gene hlyA encodes an inactive 107 kD protein (HlyA) which is activated in the cytoplasm by the hlyC gene product (HlyC). The protein products encoded by the hlyB and hlyD genes are located in the cell envelope, predominantly in the IM, and are essential for the secretion of hemolysin (Holland et al., 1986). Both hlyB and hlyD are contiguous with hlyA and are co-expressed. From DNA sequence homology studies the membrane protein HlyB has been predicted to contain a nucleotide binding site (Higgins et al., 1986).

HlyB and HlyD have been proposed to form a pore spanning both the IM and OM, allowing hemolysin to pass from the cytoplasm into the culture supernatant. However, localisation studies have suggested that HlyB and HlyD are located predominantly in the IM which is contradictory to this model (Holland et al., 1990).

Recently it has been demonstrated that TolC, a minor OM protein, is also essential for hemolysin secretion from E. coli thus indicating that the HlyB and HlyD proteins are not sufficient for secretion (Wandersman and Delepelaire, 1990). Wandersman and Delepelaire (1990) suggested that HlyB/HlyD might interact with the OM via TolC. The OM protein TolC, which is encoded by a gene which lies outside the hly gene cluster, has also been shown to share homology with prtF, a gene required for protease secretion in Erwinia chrysanthemi (Wandersman et al., 1990).

This group of hemolysin and related proteins are unique and distinct from the other major groups of proteins secreted by Gram-negative bacteria in that they contain no N-terminal signal-sequences, and secretion is secA independent (Mackman et al., 1987). The information for targeting hemolysin

was shown to be located in the last 220 amino acid residues of the C-terminal region (Mackman et al., 1987). Indeed, the C-terminal region has been used to direct a chimeric protein, containing ten amino acid residues of β-galactosidase and a large fragment of OmpF, across the cell envelope of E. coli (Mackman et al., 1987). In this latter report it was demonstrated that the last 27 amino acid residues might be sufficient to transfer this chimeric protein out of the cell. Recently, it has been demonstrated that the efficiency of hemolysin secretion is reduced when there are deletions or disruptions in the C-terminal 54 amino-acid residues. The 54 residue C-terminal region was shown to contain several structural features; a potential 18 amino-acid amphiphilic α-helix, a cluster of charged residues and a weak hydrophobic region containing hydroxylated residues (Koronakis et al., 1989). Work by Gilson et al. (1990) contradicts findings that the C-terminal of hemolysin-like proteins is responsible for targeting. Gilson et al. (1990) found that 39 N-terminal amino acid residues of CvaC were sufficient to target a CvaC-PhoA fusion protein across the IM into the periplasm of E. coli.

It is now known that a number of different bacteria use this one-step system to secrete different types of proteins and these are listed in Table 1.5. They include hemolysin [HlyA] from E. coli (Holland et al., 1990), colicin V [CvaC] from E. coli (Gilson et al., 1990), Actinobacillus pleuropneumoniae (Gygi et al., 1990), Morganella morganii (Koronakis et al., 1987) and Proteus vulgaris (Koronakis et al., 1988b), leukotoxin [IktA] from Pasteurella haemolytica (Strathdee and Lo, 1989), proteases [PrtB and PrtC] from Er. chrysanthemi (Letoffe et al., 1990), metalloprotease [PrtSM] from Serratia marcescens (Nakahama et al., 1986) and cyclolysin [CyaA] from Bordetella pertussis (Glaser et al., 1988). The proteases secreted by Er. chrysanthemi and Serratia spp.(PrtB, PrtC and PrtSM) do undergo a processing

Table 1.5. Proteins which are secreted from Gram-negative bacteria in a single step

Bacterium	Protein	Secretion factors	Heterologous host/ (Secretion factors)	Reference
E. coli (Ec)	hemolysin (HylA)	HylB, HylD, TolC	Not determined	Holland et al., 1990
S. marcescens (Sm)	metalloprotease (PrtSM)	Not determined	E. coli (PrtD,PrtE,PrtF from Ech)	Letoffe et al., 1991
Er. chrysanthemi (Ech)	protease B (PrtB) protease C (PrtC)	PrtD, PrtE, PrtF	Not determined	Delepelaire and Wandersman,1989,1990
Ps. aeruginosa (Pa)	alkaline protease (Apr)	Not determined	E. coli (PrtD,PrtE,PrtF from Ech)or (HylB,HylD, TolC from E. coli)	Guzzo et al.,1991
M. morganii	hemolysin (HylA)	HylB, HylD, HylX?	E. coli (HylB,HylD,TolC from E. coli)	Koronakis et al., 1989
Pa. hemolytica	leukotoxin (LktA)	LktB, LktD, LktX?	E. coli (HylB,HylD,TolC? from E. coli)	Strathdee and Lo, 1989
Pr. vulgaris	hemolysin (HylA)	HylB, HylD, HylX?	E. coli (HylB,HylD,TolC? from E. coli)	Koronakis et al., 1989
B. pertussis	cyclolysin (CyaA)	CyaB, CyaD, CyaE	E. coli (HylB,HylD,TolC from E. coli)	Blight and Holland, 1990
Actinobacillus pleuropneumoniae	hemolysin (HylA)	Not determined	E. coli (HylB,HylD,TolC from E. coli)	Gygi et al., 1990

event. However, the cleavage of an N-terminal region is necessary for the activation of the protease and not its secretion (Delepelaire and Wandersman, 1989). Furthermore, the hemolysin proteins from P. vulgaris and M. morganii can be secreted by the E. coli HlyB and HlyD proteins (Koronakis et al., 1987), even though the primary sequence of these hemolysin C-terminal regions (103 residues) is not conserved. However, similarities were identified between the predicted secondary structures of these regions. More recently, the heterologous secretion of a Pseudomonas aeruginosa alkaline protease (Apr) by secretion factors from Er. chrysanthemi (PrtD, PrtE and PrtF) and E. coli (HlyB, HlyD and TolC) has been reported (Guzzo et al., 1991). Both of these experiments were carried out in E. coli. It was demonstrated that Apr was more efficiently secreted by the Er. chrysanthemi pro-proteins than the E. coli Hly proteins. In a separate study, the secretion of Serratia marcescens metalloprotease by E. coli in the presence of Er. chrysanthemi protease secretion functions was demonstrated (Letoffe et al., 1991).

These findings show a remarkable degree of interchangeability exists between the one-step secretion systems of Er. chrysanthemi, Serratia marcescens, E. coli, Ps. aeruginosa, Pr. vulgaris, B. pertussis and M. morganii, although all the possible combinations have yet to be attempted.

Recently, it has been demonstrated that the Rhizobium leguminosarum nodulation gene nodO product (NodO) is similar to the 'HlyA-like' proteins (HlyA from E. coli, PrtSM from Serratia sp. and CyaA from B. pertussis)(Economou et al., 1990). The similarity was particularly strong at the N-terminal regions of these proteins where a multiple tandem repeat of a nine amino acid domain was identified. This domain, which occurs in the family of 'HlyA-like' proteins, is proposed to be a putative Ca^{2+} binding domain. NodO is thought to be involved in the early stages of nodulation and

might interact with plant root cells in a Ca^{2+} dependent manner.

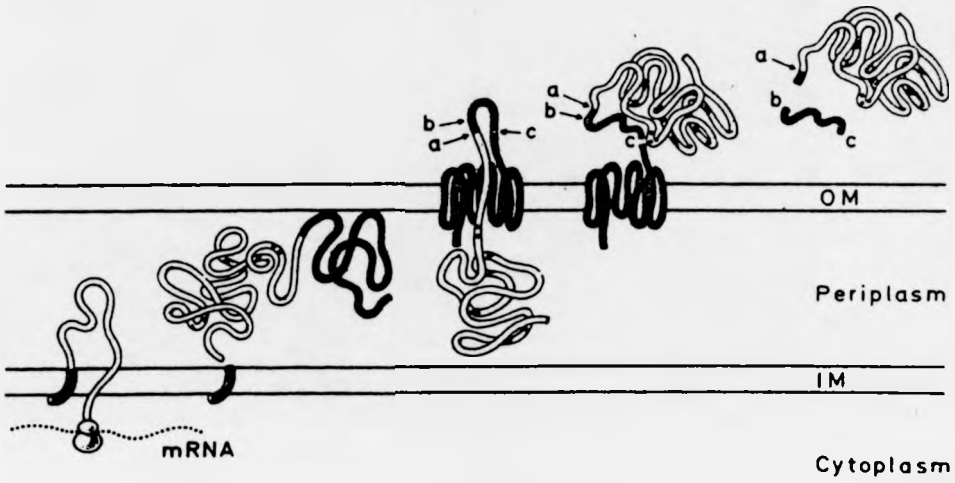
A new family of hemolysin proteins have recently been discovered. The hemolysin toxins from Serratia marcescens (ShlA) and Proteus mirabilis (HpmA) differ from the HlyA-like group in at least two ways. First, they are synthesised with N-terminal signal-sequences which are processed during secretion. Second, only one other protein (ShlB or HpmB respectively), which is located in the OM, is thought to be necessary for secretion across this membrane (Uphoff and Welch, 1990).

1.7.5. Secretion by autocatalysis

The secretion of IgA protease from Neisseria gonorrhoeae is proposed to consist of several steps (Pohlner et al., 1987). IgA protease is initially synthesised as a precursor with an N-terminal signal-sequence which directs the protein across the IM and into the periplasm. The C-terminal region of the protein, which is rich in amphipathic β-sheets, is then thought to interact with the OM. This interaction is proposed to result in the formation of a pore in the OM through which the rest of the protein passes. The extracellular IgA protease is then further processed by an autocatalytic mechanism to yield the mature IgA protease. The sequence of events is illustrated in Figure 1.6. More recently, hybrid proteins between the C-terminal region of IgA protease and Cholera toxin B subunit (CtxB) have been constructed (Klauser et al., 1990). When expressed in Salmonella typhimurium the IgA/Ctx hybrid protein was located in the OM with the CtxB moiety on the outerface of this membrane.

Serine protease from S. marcescens is another example of a protein which is secreted by this mechanism (Miyazaki et al., 1989).

Figure 1.6. Secretion of IgA protease from Neisseria gonorrhoeae



The signal-sequence directs the IgA protease precursor to the periplasm. The carboxy-terminal helper then interacts with the OM and forms a pore. The protease domain then passes through the pore and is released (from the helper) by autoproteolysis. This results in the mature form of the enzyme.

From Pohlner et al. (1987).

1.7.6. The two-step secretion mechanism

This secretion system is commonly used by Gram-negative bacteria for directing a number of different proteins across the cell wall. According to this model, proteins to be secreted are initially synthesised as precursors with N-terminal signal-sequences. The signal-sequence is sufficient for exporting the protein across the IM and into the periplasm. During the export process the signal-sequence is processed by the general export machinery (section 1.4.2.). Once in the periplasm, a further step (secretion) is necessary to transport the protein across the OM to the extracellular environment. The second step, transport across the OM, usually proceeds without any further processing of the translocated protein. Evidence for the existence of a two-step secretion mechanism is supported by findings in several laboratories and will be discussed below.

First, mutants have been generated which are defective in extracellular enzyme production. However, these extracellular enzymes are still synthesised but remain cell associated, predominantly in the periplasm. Some examples of this class of mutant are given in Table 1.6. These mutants are proposed to be unaffected in export but blocked in the secretion step. Mutants of this type are often pleiotropic in that two or more different classes of enzymes are affected in secretion. Interestingly, the same mutants are often unaffected in the secretion of other (often hemolysin/protease-like) classes of extracellular proteins. These observations suggest that different classes of enzymes share a common secretion mechanism and that some bacteria have more than one protein trafficking system. The best studied of this class are the Out- mutants of Erwinia species. This class of mutant will be discussed in greater detail in the results sections of this thesis.

Second, when extracellular enzyme structural genes from a range of

Table 1.6. Mutations which result in the periplasmic accumulation of normally secreted proteins

Bacterium	Mutant designation	Enzymes accumulated	Site of accumulation	Enzymes unnafected	References
Erwinia sp.	Out-	Pectinase Cellulase	Periplasm Periplasm	Protease	Andro *et al.*, 1984; Thurn & Chatterjee, 1985; This study
Aeromonas hydrophila	-	Aerolysin Protease Amylase	Periplasm Periplasm Periplasm	-	Jiang & Howard, 1991
Xanthomonas campestris	Xps- (formerly Xex-)	Cellulase Amylase Polygalacturonase Protease	ND ND Periplasm ND	-	Dow *et al.*, 1987; Dums *et al.*, 1991
Pseudomonas aeruginosa	Xcp-	Exotoxin Elastase Alkaline-phosphatase Phosholipase C Lipase	None Periplasmic Periplasmic Periplasmic Periplasmic	Alkaline-protease	Wretlind & Pavlovskis, 1984; Bally *et al.*, 1989, 1991

Key: ND = not determined
- = not known

Gram-negative bacteria are expressed in E. coli, the enzymes accumulate within the periplasm. Some examples are given in Table 1.7. These findings demonstrate that the E. coli export apparatus can recognise and process extracellular enzymes from a variety of different bacteria and direct them across the IM into the periplasm. However, E. coli is incapable of completing the secretion process and this inability results in an accumulation of the foreign enzymes in the periplasm. E. coli can secrete some of the proteins listed in Table 1.7. when genes encoding other (secretion) factors are also present. The secretion of pullulanase from E. coli is dependent upon the products of a further 14 K. oxytoca genes (Pugsley et al., 1990a). Also, pectate lyase from Er. chrysanthemi is only secreted by E. coli in the presence of a cosmid encoding secretion functions (He et al., 1991a). He et al. also demonstrated that the secretion of Pel from E. coli was dependent upon the general export machinery (He et al., 1991b). The secretion of pullulanase from K. oxytoca and E. coli will be discussed in section 1.7.6.1.

Finally, a periplasmic intermediate of some secreted proteins has been detected. The secretion of enterotoxin from Vibrio cholerae is particularly interesting. Toxin subunits (A subunit [28 kD] and B subunit [12 kD]) are first exported into the periplasm. They then assemble into the holotoxin (one A subunit: five B subunits) prior to translocation across the OM (Hirst and Holmgren, 1987). This result suggests that the translocation of holotoxin across the OM proceeds by a different mechanism than the export step across the IM. Protein export requires a translocation competent (relaxed) substrate as described in section 1.4.2.3. However, the holotoxin of V. cholerae traverses the OM as a mature molecule with tertiary and quaternary structure. Similar findings have been reported by Pugsley et al.

Table 1.7. Location of extracellular enzymes from various Gram-negative bacteria when expressed in *E. coli*

Extracellular enzyme	Origin of gene	Location in *E. coli*	Reference
Pectate lyase	*Erwinia* sp.	periplasm	Hinton *et al.*, 1989a Keen *et al.*,1984
Endopolygalacturonase		periplasm	Saarilahti *et al.*, 1990a
Cellulase			V.Cooper, Warwick, pers.comm
Exopolygalacuronosidase		periplasm	He and Collmer, 1990
Pectate lyase	*Yersinia pseudotuberculosis*	periplasm	Manulis *et al.*, 1988
Pullulanase	*Klebsiella oxytoca*	periplasm	d'Enfert and Pugsley, 1987
Amylase	*Aeromonas hydrophila*	periplasm	Gobius and Pemberton, 1988
DNase	*Vibrio cholerae*	periplasm	Focareta and Manning, 1987
Aerolysin		periplasm	Howard and Buckley, 1986
Cholera toxin		periplasm	Pearson and Mekalanos, 1982
Polygalacturonase	*Pseudomonas solanacearum*	periplasm	Huang and Schell, 1990b
Endoglucanase		periplasm	Huang and Schell, 1990a
Exotoxin A	*Pseudomonas aeruginosa*	periplasm	Lory *et al.*, 1988 Douglas *et al.*, 1987

(1991a and b) who investigated the secretion of pullulanase from E. coli. Pugsley et al. (1991a and b) demonstrated the existence of an intermediate of pullulanase during its secretion. The pullulanase intermediate had a relatively ordered structure (resistant to protease attack) during its translocation across the OM, unlike the cytoplasmic form (export competent) which was less ordered.

Studies using Pseudomonas solanacearum have suggested that polygalacturonase (PglA) is secreted by a two-step process involving a periplasmic intermediate (Huang and Schell, 1990a). Another extracellular enzyme produced by this bacterium, endoglucanase (EGL), is also secreted by a two-step mechanism (Huang and Schell,1990b). This enzyme is processed during its export across the IM but does not form a periplasmic intermediate (Huang and Schell,1990b). Furthermore, EGL undergoes a further processing step during its translation across the OM. This suggests that P. solanacearum might have two distinct two-step secretion pathways.

The evidence presented above suggests that a two-step mechanism, with a periplasmic intermediate, occurs for some secreted proteins. Although secretion mutants accumulate normally extracellular proteins within the periplasm, this might not represent what happens in the wild-type secreting strain. An alternative explanation might be that extracellular enzymes, when expressed in secretion mutants or in the functionally-equivalent E. coli, might be re-routed into the periplasm as a consequence of not being able to traverse the OM. This could be to prevent the jamming of the general export pathway which, as described in section 1.4.2., is a lethal event. The finding that pleiotropic secretion mutants accumulate extracellular enzymes within the periplasmic compartment could be purely artefactual and may misrepresent the actual secretion events in the wild-type strain. However, all

of the evidence so far suggests that some proteins are secreted by a two-step mechanism and proceed via a transient periplasmic intermediate.

The secretion of pullulanase is the best characterised example of two step secretion and merits special attention. The major findings from the laboratory of A. Pugsley are summarised below.

1.7.6.1. The secretion of pullulanase by K. oxytoca and E. coli

The secretion of pullulanase by *K. oxytoca* has recently been reviewed (Pugsley *et al.*, 1990a). Pullulanase is a starch debranching lipoprotein and is induced by maltose (Kornacker and Pugsley, 1989). The pullulanase structural gene (*pulA*) has been introduced into *E. coli*. The *pulA* gene is expressed and its product, PulA, is localised in the periplasm of *E. coli*. The products of a further 14 genes are required for the complete secretion of PulA by *E. coli* (Pugsley *et al.*, 1990a; d'Enfert and Pugsley, 1989; Pugsley and Reyss, 1990; Reyss and Pugsley, 1990). Pullulanase is secreted to the extracellular medium by two distinct steps. The first step involves the general export pathway (GEP) and results in the export of pullulanase across the IM into the periplasm (Pugsley *et al.*,1990b, 1991a). PulA is processed during its transfer across the IM by lipoprotein signal peptidase and fatty acylated (Pugsley *et al.*, 1990a). The second step in the secretion of PulA is its transfer across the OM and requires the 14 genes mentioned previously (Pugsley *et al.*, 1990a). The exact functions of the *pul* secretion genes are not known. All the Pul proteins encoded by the *pul* genes, apart from PulE, PulF, PulI and PulB, have been identified by either immunoblotting, *lacZ* promoter control, *E. coli* minicells or T7 gene 10 promoter expression (Pugsley *et al.*, 1990a). All of the Pul proteins identified (PulC, PulG, PulH, PulJ, PulK, PulM, PulN and PulO) were located in the IM except PulE, which was cytoplasmic and

both PulD and PulS, which were located in the OM (Pugsley et al., 1990a; d'Enfert and Pugsley, 1989; Pugsley and Reyss, 1990; Reyss and Pugsley, 1990). Gene fusions between pulA and phoA in E. coli (containing the necessary Pul secretion factors) enabled the hybrid proteins (PulA-PhoA) to be directed to the OM of E. coli but never fully secreted (d'Enfert and Pugsley, 1987). However, fusions between the C-terminal encoding region of pulA and the gene encoding β-lactamase (bla) resulted in hybrid proteins which were fully secreted by E. coli containing the Pul secretion factors (Kornacker and Pugsley, 1990). This suggests that the Pul secretion machinery can be used to secrete normally periplasmic proteins (Kornacker and Pugsley, 1990)

Some of the components of the Pul secretion machinery will be discussed in greater detail, and compared with findings from this study in section 6 of this thesis.

1.7.7. Protein conformation during translocation across the Gram-negative OM

The possible conformation of proteins traversing the IM of E. coli during export (section 1.4.2.3.) or traversing the ER membrane (section 1.3.2.) have been discussed. There is general agreement that proteins are translocated as relaxed molecules and that these conformations are ensured by various mechanisms (sections 1.3.2. and 1.4.2.3.). In the previous section (see Table 1.6.) evidence was presented which supports the existence of periplasmic intermediates for some classes of secreted proteins. The conformation that these proteins transiently attain whilst in the periplasm is important in the context of their subsequent translocation across the OM. Do these proteins fold to form secondary structures or remain relaxed prior to traversing the OM? If they remain relaxed, do they require stabilising factors

(chaperones) as do pre-proteins prior to export, or are they unfolded prior to translocation? Alternatively, can proteins with significant secondary structure traverse the OM?

Work carried out by Hirst and co-workers also suggests that assembled holotoxin produced by V. cholerae traverses the OM (see section 1.7.6.). Furthermore, pullulanase has also been shown to adopt significant secondary structure prior to translocation across the OM (section 1.7.6.). These observations suggests that proteins with significant secondary structure can traverse the OM. If so, the way that proteins cross the OM might be very different to the way they traverse the IM.

1.7.8. Assembly of OM appendages

The formation of proteinaceous appendages such as pili (the term pili is interchangeable with fimbriae) and flagella might be considered to constitute a form of 'incomplete' secretion. Flagella and pili are composed of flagellin and pilin subunits respectively. Pilin and flagellin monomers are targeted to the bacterial cell surface where they are assembled into pili and flagella and remain cell-associated (Lindberg et al., 1989).

Flagella are involved in the motility of bacteria (Driks et al., 1989). Flagellin monomers are not synthesised as precursors with N-terminal signal-sequences (Zieg and Simon, 1980). The biogenesis of functional flagella is a complex process. Flagellin subunits are polymerised at the tip of a hook which is embedded in the cell envelope (Pugsley and Schwartz, 1985). Flagellin subunits then pass through the growing flagellum and are polymerised at the tip. Mutants have been made which secrete flagellin into the growth medium. These mutants were defective in hook associated proteins (HAP) and unable to initiate polymerisation (Homma et al., 1984).

Pili are involved in a number of processes including adhesion, competence, conjugation and motility (Ottow, 1975). Pili have been grouped into six groups on the basis of morphology and function (Ottow, 1975). The molecular basis for pilus assembly has been studied for P, type1 and K88 pili (Lindberg et al., 1989). Two accessory proteins involved in P, type1 and K88 pilus assembly have been identified. One of these proteins (80-90 kD) is located in the OM and might act as an assembly platform for the pilus. The other protein (28 kD) is thought to be involved in stabilising pilin subunits prior to polymerisation (Lindberg et al., 1989). The 28 kD protein involved in P pilus assembly is proposed to be a periplasmic chaperone (Lindberg et al., 1989; Sambrook and Gething, 1989; Holmgren and Branden, 1989).

1.8. Protein targeting across the Gram-negative bacterial OM

An immense amount of research investigating the mechanisms of protein export across the E. coli IM has taken place. The major components of the export apparatus have been defined and work is now focussing on determining the roles of these individual components in protein export. At the onset of this study little was known about the translocation of proteins across the Gram-negative OM. This was probably partly due to the fact that E. coli (the most intensely studied Gram-negative bacterium) secretes very few extracellular proteins. In this laboratory we chose to investigate the Gram-negative bacterium Erwinia carotovora subsp. carotovora. This bacterium is a prolific producer of extracellular enzymes. The rest of this introduction will be devoted to the erwinias and to support the case for investigating protein targeting in Erwinia carotovora subsp. carotovora.

1.9. Protein targeting in Erwinia carotovora subsp. carotovora

1.9.1. Taxonomy of Erwinia spp.

The erwinias are phytopathogenic, Gram-negative, non-spore-forming, facultatively anaerobic bacilli, belonging to the Enterobacteriaceae family (Perombelon, 1987). A characteristic of the erwinias is the production of vast quantities of extracellular plant cell wall degrading enzymes, particularly pectolytic enzymes (Perombelon, 1987). Examples of species belonging to the Erwinia genus and the diseases they are associated with are given in Table 1.8.

1.9.2. Commercial applications of Erwinia spp.

The erwinias are important in various areas of biotechnology. These include enzyme production, antibiotic production and vitamin production. Some of these are shown in Table 1.9. The most important application of Erwinia spp.to date is the production of the anti-leukaemic drug L-asparaginase by Erwinia chrysanthemi NCPPB1066. L-asparaginase causes a depletion in circulatory L-asparagine to which lymphatic carcinomas are sensitive (Gilbert et al., 1986).

It might also be possible to exploit the plant tissue macerating enzymes produced by soft rot Erwinia spp. The use of purified Pme (free from pectin/pectate depolymerising enzymes) in the food industry has been suggested (Plastow, 1988).

1.9.3. The erwinias are important phytopathogens

Some of the diseases caused by erwinias have been listed (Table 1.8.). Within the Erwinia genus are a particularly important group which have been

Table 1.8. Diseases caused by phytopathogenic Erwinia species

Erwinia species	Disease/Host	Reference
The soft rot group		
Erwinia chrysanthemi Erwinia carotovora ssp. carotovora Erwinia carotovora ssp. atroseptica	Blackleg and soft rot in a number of different crops including potato, carrot, green pepper, chicory, celery and cucumber. Worldwide crop losses were >$100 million in 1980.	Hinton et al., 1989a; Perombelon & Kelman,1987; Kotoujansky,1987
Erwinia cypripedii	Brown rot in Cypripedium orchids	Lelliot & Dickey, 1984
Erwinia rhapontici	Soft rot in rhubarb, wheat and hyacinth	Lelliot & Dickey, 1984
Erwinia cactii	Soft rot in cacti	Alcorn et al., 1991
Erwinia amylovora	Fireblight on rosaceous plants e.g. pears	Barny et al., 1990
Erwinia rubrifaciens	Phloem necrocis in Persian walnut trees	Lelliot & Dickey, 1984
Erwinia stewartii	Vascular wilting and blight of corn	Coplin et al., 1986
Erwinia nigrifluens	Bark canker disease in Persian walnuts	Zeitoun & Wilson, 1966
Erwinia ananas	Brown spot disease on honeydew melons	Wells et al., 1987

Table 1.9. Commercial applications of Erwinia spp.

Erwinia species	Product/property	Application/use	Reference
Erwinia chrysanthemi Erwinia carotovora	L-asparaginase	Treatment of acute lymphoblastic leukaemia	Gilbert et al., 1986
Erwinia citreus	2-keto-L-gluconate	Precursor in synthesis of vitamin C	Grindley et al., 1988
Erwinia carotovora	Ribavirin	Broad spectrum anti-viral agent	Shirae et al.,1988
	β-lactam antibiotics	Broad spectrum anti-bacterial compound	Parker et al., 1982
Erwinia carotovora ssp. atroseptica	Flavouring alcohols	Food industry	Spinnler & Djian, 1991
Erwinia herbicola	β-lactam antibiotics	Broad spectrum anti-bacterial compound	Parker et al., 1982
	Herbocolins	Anti-fungal compound	Ishimaru et al.,1988
Erwinia ananas	Ice-nucleation	Food texturing using ice-nucleation properties	Arai & Watanabe, 1986
Erwinia uredovora	β-carotene	Anti-cancer agent	Misawa et al., 1991

termed the soft rot erwinias. This group comprises two main species; Erwinia chrysanthemi (Ech) and Erwinia carotovora (Ec). The latter species (Ec) has been further divided into three subspecies (Kotoujansky, 1987). These are carotovora (Ecc), atroseptica (Eca) and betavasculorum (Ecb). A new member belonging to the soft rot group of erwinias has recently been described. This new species, Erwinia cacticida, causes soft rot disease in cacti (Alcorn et al., 1991). Two other species belonging to the soft rot group are Erwinia cypripedii (Ecy) and Erwinia rhapontici (Erh) (Perombelon and Kelman, 1987). The soft rot erwinias, Ecy and Erh apart, are able to cause disease in a wide variety of plants as shown in Table 1.8. The last two species (Ecy and Erh) have a restricted host range (Table 1.8.). The diseases caused by erwinias in stored crops are considered to be opportunistic in nature. Figure 1.7. shows a potato which had been inoculated with Ecc and incubated in damp, anaerobic conditions.

The soft rot erwinias are closely related (biochemically and serologically) but have differing growth temperature optima and geographical distributions (Perombelon and Kelman, 1980). Eca (growth optima = <25°C) is associated with potatoes in cool climates, whereas Ech (growth up to 39°C) is associated with tropical, subtropical crops and greenhouse crops in temperate regions. Ecc (growth up to 36°C) is pathogenic in both temperate and tropical zones. Ecc is distributed worldwide and can be isolated from a number of sources including river water, rain water, soil, and air in agricultural regions. However, Ecc is not thought to be an indigenous inhabitant of the above niches but a plant derived contaminant (Orvos et al., 1990).

Figure 1.7. Potato tuber damage caused by Ecc.

This photograph shows a potato which has been inoculated with Ecc. This level of rot is typical of a tuber which has been incubated for 3-5 days after inoculation in damp, anaerobic conditions.

1.9.4. The structure of the plant cell wall

The components of the plant cell wall are shown in Figure 1.8. The major structural regions are the middle lamella, the primary wall and the secondary wall. Pectin polymers are a major constituent of the middle lamella and the primary wall. Pectin polymers are associated with other constituents of the cell wall including cellulose molecules and maybe glycoproteins. Enzymes which attack the components of the cell wall are, therefore, likely to destroy plant tissues. The soft rot erwinias produce pectolytic, cellulolytic and proteolytic enzymes. The nature of the plant degrading extracellular enzymes and their roles in plant tissue maceration will be summarised below.

1.9.5. Plant macerating enzymes produced by the soft rot erwinias

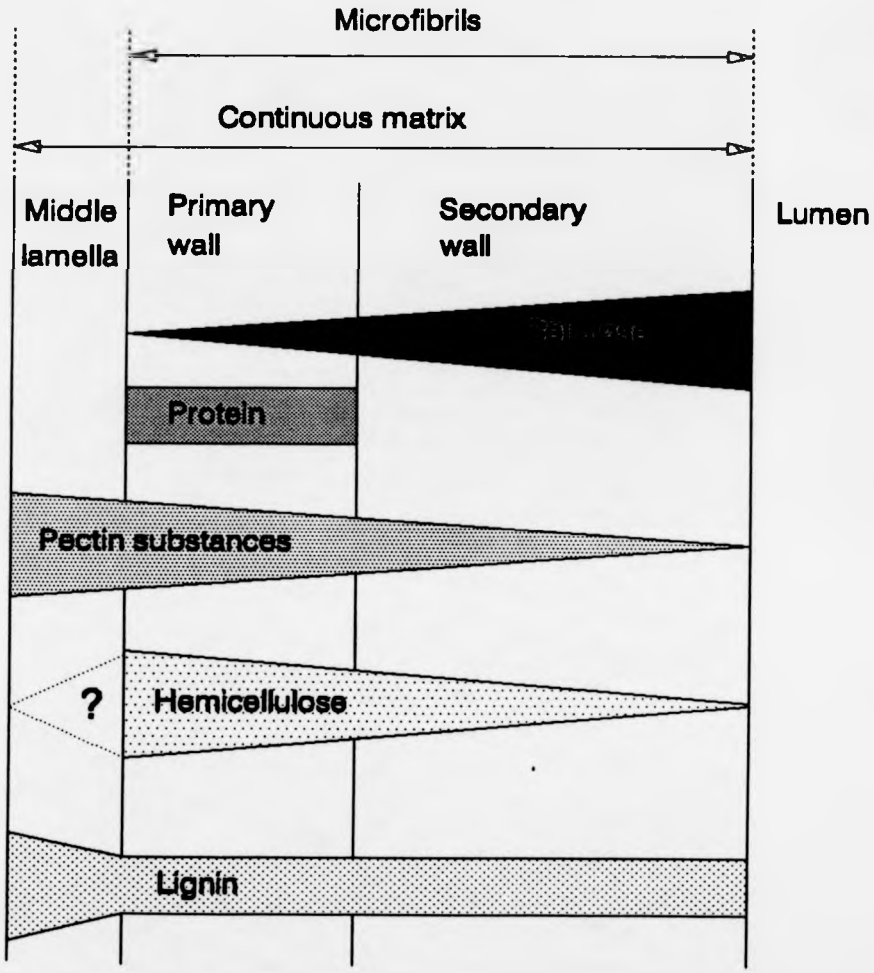
The plant tissue macerating enzymes produced by the soft rot erwinias are shown in Table 1.10. Hemicellulases have also been reported to be produced by the soft rot erwinias; however these enzymes have not been characterised. The biology of the extracellular enzymes classes will be discussed below.

1.9.5.1. The pectinases

Pectin is composed of methylated, α-1,4-linked galacturonic acid monomers (Reverchon and Robert-Baudouy 1987). The extent of methylation of this molecule is variable. The degradation of pectate and the assimilation of the breakdown products is illustrated in Figure 1.9. Figure 1.10 shows the mode of attack against pectin by the major pectolytic enzymes.

Pectin lyase (Pnl) degrades methylated pectin or methylesterified polygalacturonate (Link pectin) by trans-eliminative cleavage (Tsuyumu and

Figure 1.8. The plant cell wall



From Bateman and Basham (1976).

Table 1.10. Plant cell wall degrading enzymes produced by soft rot erwinias

Enzyme	Abb.	Location	Preferred substrate	Mode of attack
pectin methyl-esterase	Pme	extracellular	methylated galacturonan	random (endo)
pectin lyase	Pnl	extracellular	methylated galacturonan	random (endo)
endo-poly-galacturonase	endo-Peh	extracellular	demethylated galacturonan	random (endo)
exo-poly-galacturonase	exo-Peh	extracellular	demethylated galacturonan	terminal (exo)
endo-pectate lyase	endo-Pel	extracellular	demethylated galacturonan	terminal (exo)
exo-pectate-lyase	exo-Pel	cell-bound	demethylated galacturonan	terminal (exo)
cellulase[a]	Cel	extracellular	none-crystaline cellulose	random (endo)
protease[b]	Prt	extracellular	not known	not known
hemicellulase[c]				

This Table was adapted from Hinton (1986) with the following changes;

a from Kotoujansky (1987), Erwinia spp. cellulases are endo β,1-4-glucanases:
b from Wandersman (1989), three proteases have been isolated from Ech and two of these were found to be metallo-proteases:
c from Kotoujansky (1987), several arabanases and xylanases were produced by strains of Ech isolated from monocot hosts but not dicots.

Legend for Figure 1.9.

This diagram was adapted from Chatterjee et al. (1985b).

The enzymes are; (1) Peh, (2) Pel, (3) oligogalacturonate hydrolase, (4) Ogl, (5) uronate isomerase, (6) altronate oxidoreductase, (7) altronate hydrolase, (8)4-deoxy-L-threo-5-hexodiulosonate dehydrogenase, (9) 3-deoxy-D-glycero-2,5-hexodiulsonate dehydrogenase, (10) 2-keto-3-deoxygluconate kinase, (11) 2-keto-3-deoxy-6-phosphogluconate aldolase.

The elicitation of the plant response to bacterial infection has been discussed by Collmer and Keen (1986).

DKI and KDG are thought to be the inducers of Pel synthesis (Hugouvieux-Cotte-Pattat et al., 1986).

Figure 1.9. Major degradation pathways of polygalacturonate (PGA)

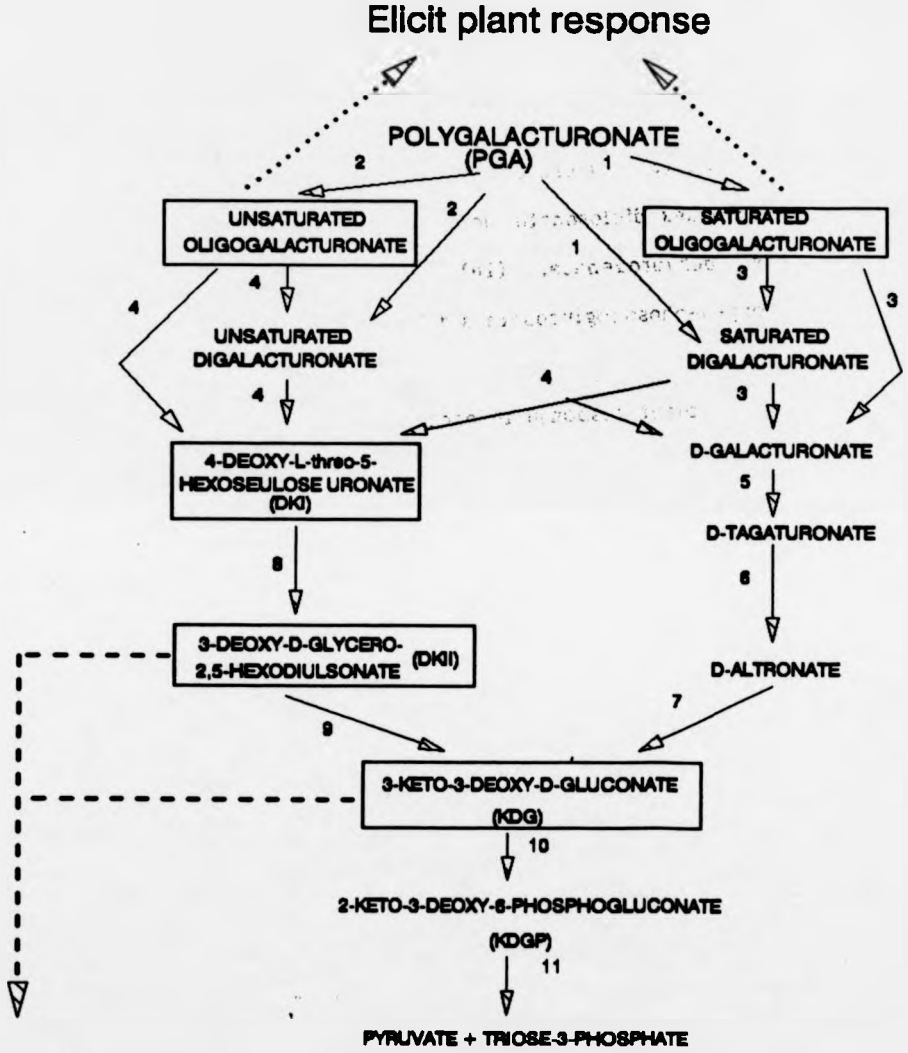
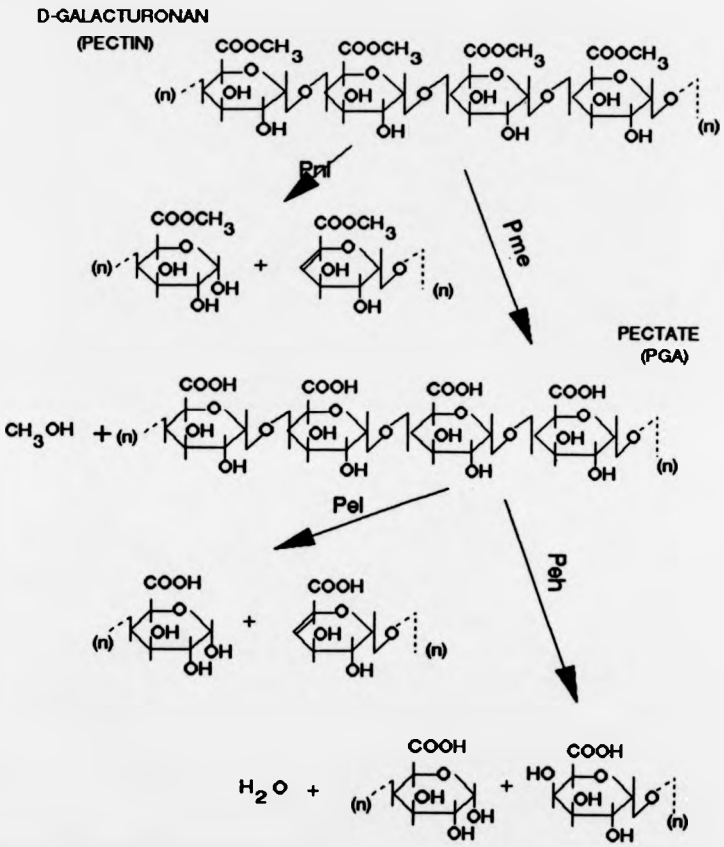


Figure 1.10. Breakdown of pectin by Pnl, Pme, Pel and Peh



Redrawn from Hinton (1986).

Chatterjee, 1984). The fate of the oligomers produced has not been reported. Pnl activity is present in culture supernatants of Ech, Ecc and Eca strains after induction with mitomycin C (Tsuyumu and Chatterjee, 1984).

An alternative pectin degradation pathway is as follows. Pectin is first de-esterified by pectin methylesterase (Pme) resulting in polygalacturonic acid (PGA) (van Gijsegem and Toussaint, 1983; Heikinheimo et al., 1991). PGA is then further degraded in two ways (Gijsegem and Toussaint, 1983). Firstly (hydrolysis), the α-1-4 glycosidic bonds of PGA are hydrolysed by endo-polygalacturonase (endo-Peh), resulting in the formation of saturated oligomeric products. These saturated oligogalacturonates are then hydrolysed into galacturonate (GA) monomers by oligogalacturonate lyase. Peh shows strict substrate specificity for pectate (unlike Pel) and has no activity against partially esterified pectin (Hinton et al., 1990). Secondly (β-elimination), PGA is attacked by pectate lyase(s), a trans-eliminative reaction producing oligomers with an unsaturated uronic acid at the non-reducing end (van Gijsegem, 1986). Unsaturated oligogalacturonates are then converted into monomers by oligogalacturonate lyase (Ogl). The monomers produced by these two pathways are then further catabolised as described in Figure 1.9.

Some of the pectin breakdown products, resulting from the activity of pectolytic enzymes, elicit plant defence responses such as the production of phytoalexins (Collmer and Keen, 1986). The main elicitors are thought to be oligogalacturonates (see Figure 1.9.) (Collmer and Keen, 1986). Further cleavage of the oligogalacturonates inactivates the activity of these elicitors (Collmer and Keen, 1986).

1.9.5.1.1. Isozymes of Pel

Isozymic forms of Pel exist in Ecc, Eca and Ech. Pel isozymes can be easily identified using flat-bed isoelectric focussing (IEF) and activity staining (Collmer et al., 1985). This technique has been used to determine the Pel isozyme profiles for several Erwinia spp. Typical isozyme profiles of Ecc and Ech are shown in Table 1.11.

1.9.5.1.2. Genetic analysis of genes encoding pectolytic enzymes

Genes encoding a variety of pectolytic enzymes have been cloned from a number of Erwinia species. Work in this area has been reviewed (Kotoujansky, 1987) and will be summarised below along with more recent findings. In most cases pectolytic genes were cloned into E. coli and recognised by the ability of the resulting colony to give a positive reaction on a pectinase detection plate (see Figure 1.11.), or by immunoscreening using anti-Pel antibodies.

Erwinia chrysanthemi strains B374 and 3937 have five pel genes arranged in two clusters. One cluster encodes PelA, PelD and PelE. The other cluster encodes PelB and PelC. Pel encoding genes have also been shown to be organised in clusters in Ecc SCRI193 (Plastow et al., 1986). The pel genes from Ecc SCRI193 have been further characterised in this laboratory (Hinton et al., 1989a). Some of the findings from this work are summarised in Table 1.11. Ecc SCRI193 makes four Pel isozymes, and localisation studies revealed that only two of the Pels, PelC and PelD, were present in the culture supernatant, whereas PelA and PelB were periplasmic. When expressed in E. coli all four Pels resided in the periplasm as would be expected (see Table 1.7.). Multiple alignment studies, using the predicted amino acid sequences from the nucleotide sequence of several pel genes, have identified

Table 1.11 Pel isozymes produced by Ech and Ecc

Bacterium	Pel isozyme	Gene symbol	Isoelectric point	Location	Group[a]	Reference
Ech (EC16) (B374)	PelB	pelB	neutral	extracellular	PLbc	Collmer and Keen, 1986
	PelC	pelC	neutral	extracellular	PLbc	Collmer and Keen, 1986
	PelA	pelA	acidic	extracellular	PLade	Collmer and Keen, 1986
	PelD	pelD	alkaline	extracellular	PLade	Collmer and Keen, 1986
	PelE	pelE	alkaline	extracellular	PLade	Collmer and Keen, 1986
Ecc (SCRI193)	PelC	pelC	alkaline	extracellular	PLbc	Hinton et al., 1989a
	PelD	pelD	alkaline	extracellular	ND	Hinton et al., 1989a
	PelA	pelA	neutral	periplasmic	ND	Hinton et al., 1989a
	PelB	pelB	neutral	periplasmic	PLy	Hinton et al., 1989a

a The groups have been determined by examining similarities in the predicted primary sequences of proteins encoded by pel genes whose DNA sequence is known and will be described further in the text.

Legend

The profiles of Pel isozymes for Ech is typical for several strains (Collmer and Keen, 1986). The Ecc strain is the one studied in this laboratory; however similar isozymes (and their localisations) have been reported for another Ecc strain, Ecc DB17 (Hinton et al., 1989a). Pels in vertical groups in the above Table (e.g. Ech PelA.D.E) were cloned on the same DNA fragment.

Table 1.12. Characterisation of genes encoding pectolytic enzymes other than Pel from Ecc and Ech

Bacterium	Strain	Enzyme	Gene	Size (kD)	Signal sequence	References
Ecc	SCRI193	endo-Peh	peh	42[b]	Yes	Hinton et al., 1990
	SCC3193	endo-PehA	pehA	42[b]	Yes	Saarilahti et al., 1990a
	193	PnlA	pnlA	32[a]	No	Chatterjee et al., 1991
Ech	EC16	exo-Peh	pehX	64[a]	Yes	He and Collmer, 1990
	B374	Pme	pme	36[b]	Yes	Plastow, 1988

a Size of protein (unprocessed) predicted from DNA sequence.

b Apparent size of mature protein from SDS PAGE.

Key

endo-Peh = endo polygalacturonase, Pnl = pectin lyase, exo-Peh = exo-polygalacturonase

Pme = pectinmethylesterase

three gene families. The PLbc family encode extracellular Pels produced by Ech, Ecc and Eca (Tamaki et al., 1988). The PLade family encode extracellular Pels from Ech alone (Tamaki et al., 1988). The PLy family encode intracellular Pels from Yersinia pseudotuberculosis, Ecc EC153 and Ecc SCRI193 (Hinton et al., 1989a). The genes encoding the PelB and PelC were sequenced and using homology studies were found to belong to previously described families of Pels. (Tamaki et al., 1988; Trollinger et al., 1989). PelC shared homology level to a family of extracellular Pels produced by Ecc, Eca and Ech. PelB is related to an intracellular family of Pels produced by Ecc and Yersinia pseudotuberculosis.

Twelve different Erwinia pel genes have been sequenced by a number of research groups (Hinton et al., 1989a). The predicted protein sequences revealed the presence of a potential signal-sequence in all the Pels studied. When all 12 Pels were aligned there were similarities running throughout the proteins. The significance of this rudimentary consensus remains unclear but there is a possibility that it may be involved in secretion or catalytic activity of Pel (Hinton et al., 1989a).

Extracellular enzymes other than Pel which are involved in the degradation of pectin are discussed below. The genes encoding such enzymes are summarised in Table 1.12. Genes encoding endo-Peh have been cloned from Ecc (Lei et al., 1985; Zinc and Chatterjee, 1985; Plastow et al., 1986; Willis et al., 1987 and Saarilahti et al., 1990a). Two of the peh genes have been sequenced (Hinton et al., 1990; Saarilahti et al., 1990a) and have been shown to encode predicted proteins of 42 kD. Both of these endo-Peh enzymes showed amino acid sequence similarity with a Peh from tomato. Endo-Peh activity has not been detected in Ech (Kotoujansky, 1987).

The gene encoding exo-Peh (pehX) was cloned from Ech and encoded a

pre-protein of M_w 64 kD (He and Collmer, 1990). No sequence similarities were found when this protein was screened against a protein data bank.

A gene encoding Pme was cloned from Ech B374 (Plastow, 1988; Heikinheimo et al., 1991) and sequenced (Plastow, 1988). The predicted pre-protein (M_w 39 kD) has a putative signal-sequence. The apparent M_w of Pme determined from SDS PAGE was found to be 36 kD.

The gene encoding Pnl was cloned from Ecc 71 (McEvoy et al., 1990). From the DNA sequence a protein of M_w 32 kD was predicted (Chatterjee et al., 1991). No putative signal-sequence was identified and N-terminal protein sequencing confirmed that Pnl was not processed during its secretion.

1.9.5.1.3. Regulation of expression of pectolytic enzymes

Pel is produced at a high basal level in the soft rot Erwinia species (Kotoujansky, 1987). Pel synthesis is induced by the breakdown products of oligogalacturonates, DKII and KDG (Hugouvieux-Cotte-Pattat et al., 1986), (see also Figure 1.9.), and is subject to catabolite repression by glucose (Kotoujansky, 1987). Pel synthesis is also repressed when levels of unsaturated digalacturonate are high enough to support bacterial growth (Collmer and Bateman, 1981; Tsuyumu, 1979). This is known as self-catabolite repression. The synthesis of Pel is also growth phase dependant with higher levels of production as the bacterial culture enters stationary phase (Kotoujansky, 1987).

Peh is constitutively produced in Ecc and is not induced by the addition of polygalacturonate (Chatterjee et al., 1981). Ech produces lower amounts of Peh but is induced (two-fold) with polygalacturonate (Chatterjee et al., 1981).

Pme is inducible but the regulation has not been studied in detail

(Hinton, 1986).

The regulation of Pnl synthesis is particularly interesting. In Ecc Pnl synthesis is induced by DNA damaging agents such as mitomycin C, UV light or nalidixic acid (McEvoy et al., 1990). Pnl induction was shown to be RecA dependent but was not induced in a rec^+ E. coli strain. It was later proposed that a positive regulator (controlled by RecA) might exist in Ecc (Chatterjee et al., 1991). The possibility that DNA damaging agents present in plants might be involved in the induction of Pnl was also suggested (McEvoy et al., 1990).

1.9.6. Cellulase production by the soft rot erwinias

The soft rot erwinias produce extracellular endo-β-1,4-glucanases which have cellulolytic activities (Kotoujansky, 1987). Endoglucanases hydrolyse glycosidic bonds in amorphous regions of cellulose in a random manner (Chambost et al., 1985). The cellulase produced by Ech 3665 is active against carboxymethylcellulose (CMC) and amorphous cellulose but not crystaline cellulose (Boyer et al., 1984 I and II). The major products of this degradation are cellobiose and cellotriose. Cellobiose is transported into the cell via the phosphotransferase system (Barras et al., 1984). Cellobiose is then further degraded by β-glucosidase.

Ech strains 3665 and 3937 each produce two biochemically and immunologically distinct cellulases. These are CelZ (encoded by celZ) and CelY (encoded by celY). The celZ and celY genes were cloned into E. coli from genomic libraries of Ech 3665 (Barras et al., 1984) and Ech 3937 (Kotoujansky et al., 1985). The nucleotide sequence of celZ has been determined. The predicted amino acid sequence shares similarities with endoglucanases from Bacillus subtilis and alkalophilic Bacillus spp. (Guiseppi

et al., 1988). Recent studies using hydrophobic cluster analysis (HCA) showed that CelZ is also related to cellulases of fungal and bacterial origin (Gilkes et al., 1991).

The cellulase produced by Ecc SCRI193 is constitutively expressed and is active against carboxymethylcellulose (CMC). Genes encoding cellulases from several Ecc strains have been cloned. These include cel clones from Ecc EC14 (Allen et al., 1986), celV from Ecc SCRI193 (V. Cooper, pers. comm.) and celS from Ecc SCC3193 (Saarilahti et al., 1990b). Work in this laboratory by V. Cooper led to the cloning and sequencing of the endoglucanase (CelV) from Ecc SCRI193. A plasmid containing celV+ complemented the Cel- mutants (RJP232, RJP213 and RJP214) generated from this work (V. cooper, pers. comm.). The predicted CelV protein has a M_w of 50 kD and a classical signal-sequence. When celV is expressed in E. coli CelV is exported to the periplasm (V. Cooper, pers. comm.).

A detailed analysis by Gilkes et al. (1991) (based on sequence similarities and hydrophobic cluster analysis (HCA)) has identified similarities between the catalytic domains of Cel proteins from a variety of sources including plants, fungi and bacteria. From such studies it was possible to identify families of Cels. Table 1.13. shows the position of cellulases from different erwinias within these families.

1.9.7. Protease production by the soft rot erwinias

Ecc and Ech produce extracellular proteases (Prt) which are secreted directly to the medium in a single step (see section 1.7.4.). The production of Prt by Ecc and Ech, when grown in minimal medium, is induced by polygalacturonic acid (PGA) and various proteins (Dahler et al., 1990). In Ecc SCRI193 (our laboratory strain), Prt is produced only in the presence of

Table 1.13. Properties of cellulases produced by Ech and Ecc

Bacterium	Strain	Cellulase	Size[a] (kD)	pI	Group[b]	Reference
Ecc	SCRI193	CelV	50		A	V. Cooper, pers. comm.
	SCC3193	CelS	27	5.5	H	Saarilahti et al., 1990b
Ech		CelZ	45	4.5	A	Boyer et al., 1987; Guiseppi et al., 1988
		CelY	35	8.2	D	

a Size of mature protein determined by SDS page.

b As determined by Gilkes et al., 1991 using amino acid sequence identity and hydrophobic cluster analysis (HCA) on catalytic domains.

proteins but not PGA. The genes encoding Prt have been cloned from Ech (Wandersman et al., 1987; Barras et al., 1986) and Ecc (Allen et al., 1986).

The best characterised of the Erwinia spp. proteases are those produced by Ech. The proteases from Ech are encoded by three structural genes, prtA, prtB and prtC. Cosmids carrying prtA, prtB and prtC have been created from Ech B374 (Wandersman et al., 1987) and Ech EC16 (Dahler et al., 1990). These two cosmids also encoded the secretion factors necessary for the release of PrtA, PrtB and PrtC when present in E. coli and a gene (inh) encoding a protease inhibitor (Wandersman et al., 1987; Dahler et al., 1990). Further characterisation of the Prt+ cosmid from Ech B374 identified a 5.5 kb region carrying genes encoding PrtB and PrtC, the protease inhibitor (Inh) and the secretion factors PrtD, PrtE and PrtF (Letoffe et al., 1990). See section 1.7.4. for a discussion of similar secretion systems in other Gram-negative bacteria.

The genetics of protease production and secretion by Ecc and Eca is not well characterised. However, Prt- mutants have been generated using Ecc EC14 (Allen et al., 1986) and Ecc SCRI193 (this study). Allen et al. were able to isolate a cosmid which restored the Prt+ phenotype in a Prt-mutant. In this laboratory it has been possible to complement some of the Prt- mutants which were isolated as part of this work. However, it was not possible to isolate a cosmid which would produce Prt in E. coli (P. Davies, pers. comm.).

1.9.8. Secretion of extracellular enzymes from Erwinia spp.

The secretion of extracellular Pel, Cel and Prt by Erwinia spp. is achieved without the concomitant release of other intracellular proteins (Andro et al., 1984; Kotoujansky, 1987). This suggests that Erwinia spp. have

specific secretion mechanisms in order to target the extracellular enzymes beyond the boundaries of the cell wall. Mutants have been generated in both Ecc and Ech which are unable to localise correctly some classes of extracellular enzymes. These are described in Table 1.14. The class of mutant designated Out- accumulates Pel and Cel intracellularly but is unaffected for Prt synthesis or secretion. Mutants of Ecc SCRI193 have also been generated which are affected in the production of Prt but unaffected in the production of Pel or Cel (this study and unpublished work). These facts suggest that Erwinia species have two independent secretion mechanisms. One is for the secretion of Pel and Cel, the other for the secretion of Prt. A more detailed investigation of the Pel, Cel secretion system will be given in the results sections. The findings of other groups of workers will also be discussed in greater detail in the results sections of this thesis.

1.9.9. Are the extracellular enzymes the major determinants of pathogenicity?

The extracellular enzymes Pel, Cel and Prt are all thought to play a major role in the pathogenicity of the soft rot Erwinia species (Daniels et al., 1988). The exact roles that Prt and Cel play in pathogenesis are not known. The pectolytic enzymes have been the most extensively characterised and are considered to be the most important in causing disease (Ried and Collmer, 1988). The reasons for this are as follows: first, purified Pel can macerate plant tissue; second, E. coli transformants containing highly expressed pel genes can macerate plant tissue (Ried and Collmer, 1988).

In order to determine the roles that pectolytic enzymes play in pathogenicity, mutations have been generated within specific genes encoding pectolytic enzymes. The results from such experiments suggested that in Ech 3937 and CUCPB 1237, PelA, PelD and PelE are the most important Pel

Table 1.14. Mutants of Ecc, Eca and Ech pleiotropically defective in the secretion of Pel and Cel but unaffected in the production of Prt

Bacterium	Strain	Mutant	Enzymes affected	Enzymes unaffected	Mutagenesis	References
Ech	3665	Out-	Pel, Cel	Prt	NTG, Mu7701	Andro et al., 1984
	EC16	Out-	Pel, Cel	Prt	Tn5	Thurn & Chatterjee, 1985
	3937	Out-	Pel, Cel	Prt	Mu d(Aplac)	Ji et al., 1987
Ecc	ECC193	Out-	Pel, Cel	Prt	NTG	Chatterjee et al., 1985b
	71	Out-	Pel, Cel	Prt	Tn5	Murata et al., 1990
	SCRI193	Out-	Pel, Cel	Prt	Tn5, Tn10	Unpublished
		Out	Pel, Cel	Prt	TnphoA	Hinton & Salmond, 1987
		Out-	Pel, Cel	Prt	EMS	This study
Eca	12	Out-	Pel, Cel	Prt	EMS	Murata et al., 1990

Key: Out- = phenotype leading to periplasmic accumulation of Pel and Cel but normal secretion of Prt.
Pel = pectinases, Cel = cellulase, Prt = protease
EMS = ethylmethylsulphonate, NTG = N,methyl-N'-nitro-N-nitrosoguanidine

isozymes in pathogenicity, whereas PelB and PelC are the least important (Kotoujansky, 1987; Roeder and Collmer, 1985). Furthermore, using marker exchange eviction mutagenesis, a strain of Ech EC16 was constructed which had deletions in all of the Pel structural genes (Ried and Collmer, 1988). This mutant (UM1105), although showing reduced virulence in potato tuber maceration tests, was still able to cause significant maceration in various vegetable tissues (Ried and Collmer, 1988). The residual plant macerating activity in UM1105 was attributed to the production of exo-Peh. However, in a more recent piece of work the gene encoding this enzyme (pehX) was evicted from UM1005 (He and Collmer, 1990). The resulting mutant (CUCPB5009) was still able to macerate chrysanthemum pith tissue. He and Collmer (1990) have recently identified novel pectolytic enzymes from Ecc cultures grown on chrysanthemum cell walls which they believe might account for the maceration ability of CUCPB5009 in planta.

1.9.10. Other bacterial functions involved in pathogenesis

The importance of the extracellular enzymes (particularly pectolytic enzymes) in pathogenesis has been discussed in previous sections. investigations have been undertaken to identify other pathogenic factors.

Work carried out in this laboratory involved screening colonies of Eca SCRI1043, with random Tn5 insertions, for virulence. These colonies were inoculated into living plant stems. Reduced virulence mutants (Rvi-) were initially classified into three groups; group 1 were auxotrophs; group 2 were extracellular enzyme mutants and group 3 reduced were growth rate mutants (Hinton et al., 1989b). Subsequent studies have shown that some of these mutants were also defective in motility (Mot-) and lacked flagella (V. Mullholland, pers. comm.). The further characterisation of these mutants

is underway.

The acquisition of iron is thought to be important for pathogenicity of Ech in Saintpaulia spp. plants (Enard et al., 1988). Bacteriocin resistant mutants of Ech were generated, some of which were avirulent. Further studies suggested that these mutants might be deficient in iron assimilation.

Lipopolysaccharide (LPS) mutants have been generated in Ech 3937 by selecting for resistance to bacteriophages ϕEC2 and Mu. Phage resistant mutants which lacked the O-antigen region (high-molecular-weight heterologous chain) of LPS were avirulent (Schoonejans et al., 1987). Workers in this laboratory have isolated Eca mutants resistant to Eca phages which are pleiotropically avirulent and have altered LPS profiles (I. Toth, pers. comm.).

An elaborate technique was used by Beaulieu and van Gijsegem (1990) to identify Ech 3937 genes which were induced in planta. A culture of Ech 3937 was mutagenised with Mu dIIPR3 (which has a promotorless nptI gene) and inoculated onto minimal medium plates containing kanamycin and Saintpaulia spp. extracts. Colonies resulting from this procedure must have had a Mu dIIPR3 insertion with the nptI gene under the control of a plant (extract) inducible promoter. Several different mutant classes deficient in plant-inducible genes were isolated. These included mutants defective in acidic Pel (PelA) synthesis, the galacturonate degradation pathway, and cation uptake.

1.9.11. Genetic techniques available for investigating erwinias

The taxonomic position of the erwinias is close to that of E. coli, the most extensively studied bacterium. For this reason many of the genetic techniques developed for E. coli are also applicable to the erwinias. The

applicability of these techniques is strain dependent (Ellard *et al.*, 1989). However, *Ecc* SCRI193 is particularly genetically amenable. The genetic techniques developed for *Ecc* and *Eca* are listed in Table 1.15.

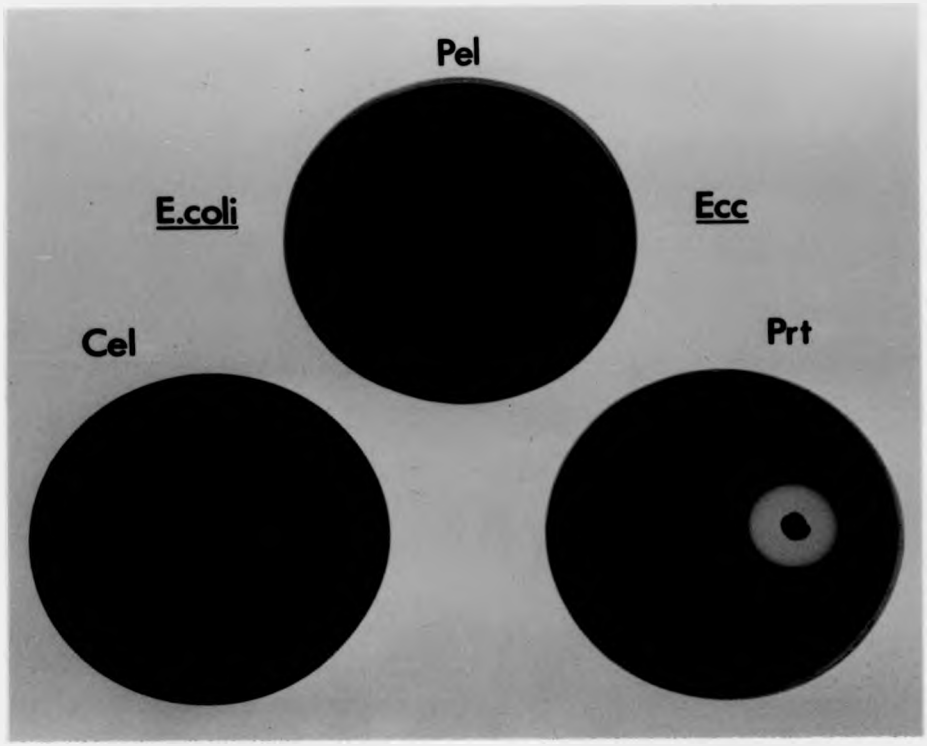
1.9.12. *Ecc* SCRI193 is an ideal bacterium for investigating protein secretion

The importance of erwinias in agriculture and industry has been discussed. Erwinias are also useful for investigating the fundamental biological process of protein secretion. Protein secretion by the erwinias is directly linked to their pathogenicity. This makes them particularly worthwhile bacteria to study.

The ability of *Ecc* to secrete a range of extracellular enzymes has been discussed. The production of the extracellular enzymes (Pel, Cel and Prt) by colonies of *Ecc* can be detected using assay plates and is shown in Figure 1.11. The halos surrounding the *Ecc* colonies represent extracellular enzyme activities on each of the three detection plates after staining. *E. coli* was inoculated alongside *Ecc* and can be seen to be deficient in the ability to produce these extracellular enzymes. The production of extracellular enzymes can also be easily quantified using spectrophotometric and biochemical assays.

Extracellular enzyme production can be investigated using a classical 'black-box' approach, thus requiring no prior knowledge of the process. Firstly, mutants are generated which exhibit a defective phenotype with respect to extracellular enzyme production. This procedure can be carried out using a number of techniques including chemical and transposon mutagenesis. The method used for generating transposon insertion mutants in *Ecc* SCRI193 is outlined in Figure 1.12. The λ delivery system can be used to introduce a range of transposons, including Tn*5* and Tn*10*, into the *Ecc* SCRI193

Figure 1.11. Extracellular enzyme production by Ecc.



Colonies of E.coli and Ecc SCRI193 were patched onto the extracellular enzyme detection plates as illustrated. These plates were incubated at 30°C for 3 days before they were developed. The halos which surround the Ecc patches indicate the production of extracellular enzyme on each of the three plates. E. coli is negative for the three extracellular enzymes tested.

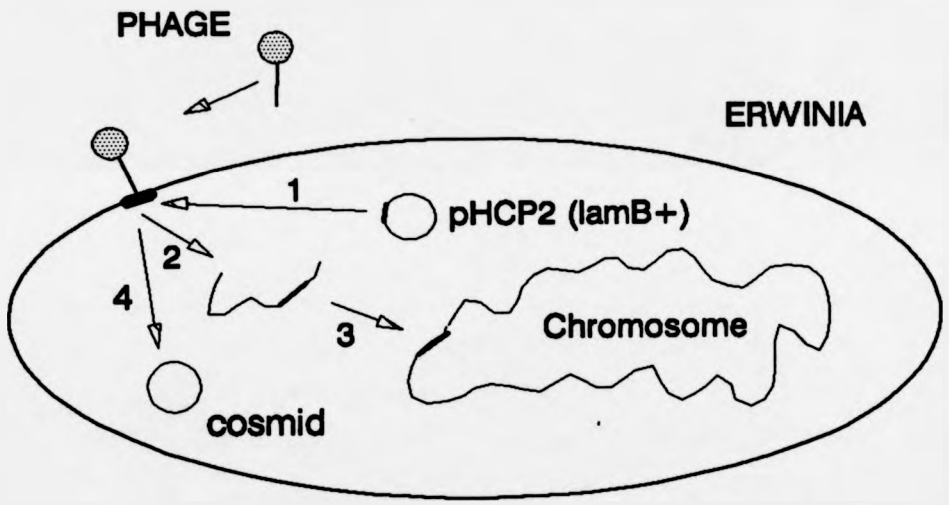
Table 1.15. Genetic systems available for Ecc & Eca

Species	Strain	System	Vector	Reference
Ecc	SCRI193	Chromosome mobilisation	R68::Mu F'lac$^+$Tc	Forbes & Perombelon, 1985
		Plasmid transformation	pBR322	Hinton et al., 1985a
		Nonsense suppressor	(pLM2)	Hinton et al., 1985b
		Transposon mutagenesis	λ	Salmond et al., 1986
		Cosmid complementation	pHC79	This study
		Electroporation	various	Unpublished
		Generalised transduction	ØKP	I. Toth Pers. comm.
		Gene fusions	TnphoA	Hinton & Salmond, 1987
	Ecc193	Chromosome Mobilisation	pULB113	Chatterjee et al., 1985a
	Ecc71	Tn5 mutagenesis	pJB4JI	Zinc et al., 1984
		Cosmid complementation	pSF6	Murata et al., 1990
	Brig-PIA	Gene fusions	pJB4JI	Jayswal et al., 1984
Eca	SCRI1043	Tn5 mutagenesis	λ	Hinton et al., 1989b
	Eca12	Chromosome mobilisation	pULB113	Chatterjee et al., 1985a

Legend for Figure 1.12.

This diagram shows an Erwinia spp. bacterium which harbours plasmid pHCP2. This plasmid carries genes encoding for Ap^r and LamB. The LamB protein correctly assembles in the Erwinia spp. OM and can serve as the receptor for bacteriophage λ (1). Phage λ can adsorb to the Erwinia spp. cell surface and inject its DNA (2). However, λ fails to replicate in this strain. Phage λ derivatives carrying transposons will inject their DNA into Erwinia spp. Transductants resulting from transposition can be selected by transposon encoded antibiotic resistance markers (3). Cosmids which have been packaged into λ heads can also be introduced into Erwinia strains via LamB (4).

Figure 1.12 Phage λ as a suicide vector for the delivery of transposons or cosmids



derivative (HC131) which carries a plasmid (pHCP2) encoding the λ receptor protein LamB.

The resulting mutants can then be investigated to determine the nature of a mutation giving rise to a particular phenotype. Using cosmid technology it is possible to complement directly mutants arising from mutagenesis procedures. Cosmids can also be introduced into this strain of Ecc using λ as a delivery vehicle. By investigating the complementing cosmids using standard molecular biology techniques it is possible to identify the genes of interest and to perform a detailed molecular analysis upon them.

1.10. Aims of this work

The aim of this project was to investigate extracellular enzyme secretion in Ecc SCRI193. The first part of this project was to generate a large number of mutants defective in extracellular enzyme secretion using the chemical mutagen EMS. This mutagen was used for the following reasons: 1) it has been used successfully before in this strain for generating a range of auxotrophic mutants (Forbes and Perombelon, 1985); 2) it should cause mutations in a random manner unlike transposons which may insert into preferential locations ('hotspots'); 3) there existed a possibility of isolating 'subtle' missense mutations including conditional mutations and 4) gross insertion mutants resulting from transposons might have been lethal.

The second part of the project was to characterise Ecc mutants from the mutagenesis procedure which were defective in extracellular enzyme secretion using physiological and genetic techniques.

CHAPTER 2

MATERIALS AND METHODS

2.1. Strains

Bacterial strains used in this study are listed in Table 2.1. Derivatives of these strains will be referred to in the text where appropriate. Plasmids used in this study are listed in Table 2.2. and derivatives are referred to in the text where appropriate. Bacteriophages are listed in Table 2.3.

2.2. Media

2.2.1. General information

Media are listed in Table 2.4. All growth media and solutions were prepared using filtered water obtained from an Elgastat Spectrum (Elga) water purification system. All media were sterilised by autoclaving at 121°C and 15 psi for 20 min. Media were solidified using 1.5% (w/v) Bacto agar. Minimal medium was supplemented with amino acids (20 µg/ml, final conc.) and sugars (0.2% (w/v), final conc.) when required. Antibiotics were prepared and stored as 100 x stocks as shown in Table 2.5. Rifampicin, tetracycline and chloramphenicol were dissolved in 50% (v/v) ethanol and stored at -20°C. Nalidixic acid was dissolved in 30 mM $NaOH$ and stored at 4°C. All other antibiotics were dissolved in sterile Elga water and stored at 4°C.

Table 2.1. Bacterial strains

Strain	Characteristics	Plasmid phenotype	Source	Reference
E.coli				
K12	Wild type		D. Gill	Bachmann & Low, 1980
K38	K12, HfrC		S. Tabor	Tabor & Richardson, 1985
DH1	F$^-$, recA1, endA1, th-1, hsdR17, gyrA96 (r_k^- m_k^-), supE44, relA1,		D. Gill	Maniatis, 1982
LE392	F-, hsdR514, (r_k^- m_k^+) supE44, supF58, lacY1, galK2, galT22, metB1 trpR55, -		J. Hinton	Maniatis, 1982
JM101	Δ(lac-pro), supE thi, F'traD36, proAB $lacI^q$, ZdeltaM15	F'	J. Hinton	Yanisch-Perron et al.,1985
TG1	JM101 ecoK derivative	F'	J. Hinton	
GJ342	R64 drd11 pLVC9	Tc^r Cm^r	D. Gill	van Haute et al., 1983
DW74	GJ342(pHCP2)	Ap^r	D. Gill	Salmond et al., 1986

Table 2.1. (cont.)

DW75	DW74(R64 drd11) (pLVC9)	Ap^r,Tc^r Cm^r	D. Gill	Salmond et al., 1986
Ecc				
SCRI193	Wild type	-	J. Hinton	Forbes &, Perombelon, 1985
HC131	SCRI193(pHCP2)	Ap^r, LamB	J. Hinton	Salmond et al., 1986
GS2001	HC131 pho::Tn10	Ap^r, LamB	J. Hinton	Hinton & Salmond,1987
AC4000	HC131 out::Tn10	Ap^r, LamB	lab stock	Unpublished
PR54	GS2001 out::Tn5	Ap^r, LamB	lab stock	Unpublished
PR33	GS2001 out::Tn5	Ap^r, LamB	lab stock	Unpublished
RJP122	HC131 out- (EMS)	Ap^r, LamB	This work	
RJP159	HC131 out- (EMS)	Ap^r, LamB	This work	
RJP190	HC131 out- (EMS)	Ap^r, LamB	This work	
RJP200	HC131 out- (EMS)	Ap^r, LamB	This work	
RJP208	HC131 out- (EMS)	Ap^r, LamB	This work	

Table 2.1. (cont.)

RJP211	HC131 out- (EMS)	Ap^r, LamB	This work
RJP220	HC131 out- (EMS)	Ap^r, LamB	This work
RJP221	HC131 out- (EMS)	Ap^r, LamB	This work
RJP233	HC131 out- (EMS)	Ap^r, LamB	This work
RJP249	HC131 out- (EMS)	Ap^r, LamB	This work
RJP250	HC131 out- (EMS)	Ap^r, LamB	This work
RJP251	HC131 out- (EMS)	Ap^r, LamB	This work
RJP253	HC131 out- (EMS)	Ap^r, LamB	This work
RJP254	HC131 out- (EMS)	Ap^r, LamB	This work

A complete list of the Ecc HC131 mutants generated by EMS mutagenesis is given in Table 3.2.

Table 2.2. Plasmids

Plasmid/ cosmid	Characteristics	Plasmid phenotype	Source	Reference
pBR322	Multicopy cloning vector	Ap^r,Tc^r	Amersham Int.	Bolivar et al., 1977
pHCP2	pBR322::lamB	Ap^r,Tc^s	J. Hinton	Clement et al.,1982
pHC79	pBR322::cos	Ap^r,Tc^r	J. Hinton	Hohn & Collins, 1980
cHIL159	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL122	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL190	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL208	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL211	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL220	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL221	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL233	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL253	pHC79::out$^+$	Ap^r,T^r	This study	
cHIL254	pHC79::out$^+$	Ap^r,T^r	This study	
pGP1-2	T7 polymerase	Kn^r	S. Tabor	Tabor & Richardson, 1985
pT7-5	T7ϕ10 promoter	Ap^r	S. Tabor	
pT7-6	T7ϕ10 promoter	Ap^r	S. Tabor	

Table 2.3. Bacteriophages

Phage	Characteristics	Source	Reference
Coliphages			
M13 mp8	Sequencing vector	Amersham Int.	Messing & Vieira, 1982
λcI857	Thermoinducible	J. Hinton	Vollenweider et al., 1980
Ecc SCRI193 bacteriophages			
ØKP	SCRI193 generalised transducing phage	I. Toth	Pers. comm.
ØD-2	SCRI193 bacteriophage	I. Toth	Pers. comm.
Ø565	SCRI193 bacteriophage	I. Toth	Pers. comm.
Ø575	SCRI193 bacteriophage	I. Toth	Pers. comm.

Table 2.4. Media

Medium	Constitutents per litre[a]
NB	13 g Oxoid nutrient broth
NBA	13 g Oxoid nutrient broth 16 g Bacto agar
2YT	15 g Bacto tryptone 10 g Bacto yeast extract 5 g NaCl
LB	10 g Bacto tryptone 5 g Bacto yeast extract 5 g NaCl [pH 7.2]
SOB	20 g Bacto tryptone 5 g Bacto yeast extract (10 ml 1 M NaCl) (2.5 ml 1 M KCL) (10 ml 1 M $MgSO_4$, 1 M $MgCl_2$ Filter Sterile) [pH 6.9 - 7.0]
SOC	as SOB (+ 20 ml 1 M glucose)
Stab medium	as NB + 7 g Bacto agar

Table 2.4. (cont.)

DDA	20 g Bacto tryptone 8 g NaCl (10 ml 1 M $MgSO_4$) 9 g Bacto agar for plates or 2.5 g Bacto agar for soft agar
Phage buffer	10 mM Tris.HCl 10 mM $MgSO_4$ 0.01% (w/v) gelatin [pH 7.4]
Minimal medium	(20 ml 50 x Phosphate) (10 ml 10% (w/v) $(NH_4)_2SO_4$) (10 ml 1% (w/v) $MgSO_4.7H_2O$) (10 ml 20% (w/v) carbon sourced)
50 x Phosphate	350 g K_2HPO_4 100 g KH_2PO_4 [pH 6.9 - 7.1]
Pectate lyase detection plates (Pel)	16 g Bacto agar (5 ml 20% (w/v) Bacto yeast extract) (10 ml 10% (w/v) $(NH_4)_2SO_4$) (1 ml 1 M $MgSO_4.7H_2O$) (9 ml 50% (w/v) glycerol) (125 ml 2% (w/v) Polygalacturonic acid) (100 ml Pel Phosphate buffer)

Table 2.4. (cont.)

Pel induction (PM) medium	(5 ml 20% (w/v) Bacto yeast extract) (10 ml 10% (w/v) $(NH_4)_2SO_4$) (1 ml 1 M $MgSO_4.7H_2O$) (9 ml 50% (w/v) glycerol) (125 ml 2% (w/v) Polygalacturonic acid) (20 ml 50 x Phosphate buffer)
Pel phosphate buffer	15 g Na_2HPO_4 anhydrous 0.7 g $NaH_2PO_4.H_2O$ [pH 8.0]
Cellulase detection plates (Cel)	10 g Sigma carboxymethyl cellulose 16 g Bacto agar (25 ml 20% (w/v) Bacto yeast extract) (4 ml 50% (w/v) glycerol) (20 ml 50 x phosphate) (10 ml 10 (w/v) $(NH_4)_2SO_4$)
Protease detection plates (Prt) (destructive)	13 g Oxoid nutrient broth 30 g Oxoid gelatin 16 g Bacto agar

Table 2.4. (cont.)

Protease detection plates (SMNA) (non-destructive)	13 g Oxoid nutrient broth (5% final (w/v) skimmed milk) 16 g Bacto agar
Freezing medium (2 x)	126 g K_2HPO_4 anhydrous 0.9 g sodium citrate 0.18 g $MgSO_4.7H_2O$ 1.8 g $(NH_4)_2SO_4$ 3.6 g KH_2PO_4 anhydrous 88 g glycerol
FCM	50% (w/v) glycerol 0.1 M $CaCl_2$ (grade 1)

[a] Items in brackets were added after autoclaving, from sterile stocks. Notes on making extracellular enzyme plates are given in section 2.2.2.

Table 2.5. Antibiotics

Antibiotic	Abbreviation	Final concentration
Sodium ampicillin	Ap	50 µg/ml
Chloramphenicol	Cm	50 µg/ml
Kanamycin sulphate	Km	50 µg/ml
Rifampicin	Rm	50 µg/ml
Spectinomycin	Sp	50 µg/ml
Streptomycin sulphate	Sm	100 µg/ml
Tetracycline	Tc	10 µg/ml

Prepared as 100 x final concentration stocks rifampicin, tetracycline and chloramphenicol were dissolved in 50% (v/v) ethanol, and stored at -20°C. All other antibiotics were dissolved in 100 ml sterile double distilled water and stored at 4°C.

Table 2.4. (cont.)

Protease detection plates (SMNA) (non-destructive)	13 g Oxoid nutrient broth (5% final (w/v) skimmed milk) 16 g Bacto agar
Freezing medium (2 x)	126 g K_2HPO_4 anhydrous 0.9 g sodium citrate 0.18 g $MgSO_4.7H_2O$ 1.8 g $(NH_4)_2SO_4$ 3.6 g KH_2PO_4 anhydrous 88 g glycerol
FCM	50% (w/v) glycerol 0.1 M $CaCl_2$ (grade 1)

[a] Items in brackets were added after autoclaving, from sterile stocks. Notes on making extracellular enzyme plates are given in section 2.2.2.

Table 2.5. Antibiotics

Antibiotic	Abbreviation	Final concentration
Sodium ampicillin	Ap	50 µg/ml
Chloramphenicol	Cm	50 µg/ml
Kanamycin sulphate	Km	50 µg/ml
Rifampicin	Rm	50 µg/ml
Spectinomycin	Sp	50 µg/ml
Streptomycin sulphate	Sm	100 µg/ml
Tetracycline	Tc	10 µg/ml

Prepared as 100 x final concentration stocks rifampicin, tetracycline and chloramphenicol were dissolved in 50% (v/v) ethanol, and stored at -20°C. All other antibiotics were dissolved in 100 ml sterile double distilled water and stored at 4°C.

2.2.2. Preparing and developing extracellular enzyme detection plates

2.2.2.1. Pectate lyase (Pel) detection plates

Pel detection plates were made according to Andro et al. (1984). Polygalacturonic acid (PGA) at 2% (w/v) was made by slowly adding 20 g Sigma PGA to 900 ml H_2O. NaOH (10 M) was then added dropwise until a pH of 5.5 was reached. The PGA then went into solution and became gelatinous and clear. The solution was then made up to 1 l and autoclaved (15 psi for 15 min) in 80 ml aliquots. The resulting PGA solution was added to water agar (cooled to 60°C) along with the other ingredients to make Pel detection plates (Table 2.2.).

Pel detection plates were developed by flooding them with a saturated solution of copper acetate. Pel+ colonies were identified by a white halo surrounding the colony against a pale blue background.

2.2.2.2. Cellulase (Cel) detection plates

Cel detection plates were modified from Gilkes et al. (1984) by J. Hinton (pers. comm.). Carboxymethylcellulose (CMC) was added to 380 ml of H_2O in a 500 ml bottle to give a concentration of 10 g per litre. Lumps of CMC were broken with a 10 ml pipette and the mixture autoclaved for 30-40 min, 15 psi until a gelatinous solution was formed. Agar was then added (1.5% w/v) before re-autoclaving the mixture for 20 min at 15 psi. The CMC/agar mix was then melted as required and cooled to 60°C before the addition of the other ingredients.

Cel plates were developed by flooding them with 0.2% (w/v) congo red solution for 20 min, then bleached with 1 M NaCl for 15 min. They were

then counter-stained with 1 M HCl for 5 min. Cel^+ colonies had clear/pink translucent halos against a dark blue background.

2.2.2.3. Protease (Prt) detection plates

Prt detection plates have been described by Thurn and Chatterjee (1985). Prt detection plates were made by dissolving gelatin in H_2O at 60°C to a concentration of 3% (w/v). NA was then added to the standard concentration (28 g per l). The medium was then sterilised by autoclaving.

Plates were developed by flooding with either acid mercuric chloride (for high quality halo definition) or routinely with a saturated ammonium acetate solution. Acidic mercuric chloride was made by mixing together 20 ml conc. HCl with 80 ml H_2O, and dissolving into this solution, with stirring, 15 g $HgCl_2$. Care was taken due to the toxicity of the solution. Prt^+ colonies gave clear zones on an opaque white background.

2.3. Reagents

Reagents for media were obtained from BDH and Fisons and were of "Analar" grade. Polygalacturonic acid, carboxymethylcellulose, amino acids and vitamins were obtained from Sigma. Skimmed milk, yeast extract (YE), nutrient agar (NA), nutrient broth (NB) and tryptone were obtained from Difco. All other chemicals were obtained from Sigma except the following. Polyethylene glycol (PEG 6000), phenol and Trizma base were purchased from BDH. Caesium chloride was obtained from BRL. Ammonium persulphate, N:N'-methylene-bis-acrylamide, and acrylamide were purchased from Bio-Rad. Enzymes for molecular biology were obtained from Bethesda Research Laboratories (BRL), Amersham International, Boehringer Mannheim and New England Biolabs (NEB). Rainbow molecular weight markers, bacteriophage

DNA, plasmids pBR322 and pHC79, ^{35}S methionine, ^{35}SdATP and the *in vitro* transcription/translation (Zubay) kit were purchased from Amersham International. 'Sequenase' was purchased from United States Biochemical Corporation (USBC).

2.4. Maintenance of strains

E. coli and *Ecc* strains were stored on NA plates containing appropriate antibiotics for up to 3 months. For long term storage all strains were frozen in duplicate at -70°C. Cultures of strains grown under selective conditions were mixed with an equal volume of 2 x freezing medium and stored at -70°C.

2.5. Growth of bacteria

E. coli and *Ecc* were incubated at 37°C and 30°C respectively unless stated otherwise. Liquid cultures were generally grown in 25 ml screw cap bottles (universal tubes) in an orbital shaker (200 rpm). For larger culture volumes, conical flasks were used containing liquid medium at no more than 1/10 total flask volume.

2.6. Chemical mutagenesis using ethyl methyl sulphonate (EMS)

The protocol was adapted from Forbes and Perombelon (1985). *Ecc* was subcultured 1:25 from a 10 ml overnight culture (grown in LB at 30°C with shaking) into 25 ml of LB (in a 250 ml flask). The culture was incubated with shaking (200 rpm) at 30°C until reaching A600 = 0.6. The viability of the culture at this stage was determined before the addition of EMS (500 µl). The viability of the culture immediately prior to the addition of EMS (t=0 min) and every 30 min thereafter was determined by serial

dilution. Samples were washed in TMG buffer (see section 2.8.) before serial dilution to remove EMS. Serial dilutions were inoculated onto NA plates containing ampicillin and incubated for 2 days at 30°C. Conditions resulting in 1-4% survival were determined and another culture was treated, using these conditions, for the generation of mutants.

2.7. Transposon mutagenesis

2.7.1 Ecc

Strains were incubated overnight at 30°C with shaking in 10 ml LB containing Ap (50 μg/ml). Cells were then pelleted and resuspended in 1 ml LB containing $MgSO_4$ (10^{-2}). A 100 μl aliquot of cells was removed as a spontaneous antibiotic control before the addition of 100 μl of phage lysate (10^{10} pfu). After 20 min static incubation at 30°C (to allow adsorption), a further 10 ml of fresh LB was added and the culture was incubated with shaking at 30°C for a further 90 min. Cells were then pelleted, resuspended in 1 ml of LB and 100 μl aliquots were plated onto NA plates containing the appropriate antibiotic and incubated for 36 hr at 30°C. This method yielded more than 2000 transductants per 10^{10} pfu of phage lysate.

2.7.2. E. coli

The above procedure was followed except that the incubation temperature throughout (apart from adsorption at RT) was 37°C, and 0.2% (w/v) maltose was present in all growth media.

2.7.3. Transposition onto plasmids

The general procedure as above was followed except that increased levels of antibiotic were used to select transductants (Kn 300 µg/ml).

2.8. Preparing high titre bacteriophage lysates

All λ suicide vectors were propagated on LE392 (the E. coli supE, supF strain). All Ecc bacteriophages were propagated on Ecc HC131 (LamB+).

Lysates were initially plaque purified by titrating bacteriophage onto sensitive bacterial lawns to produce single plaques. To do this the phage stock was serially diluted 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8} using TMG buffer (10 mM Tris.HCl (pH 7.4), 10 mM $MgSO_4$, 0.01% (w/v) gelatin). Aliquots (100 µl) were mixed with 200 µl of an overnight (stationary phase) culture of a sensitive bacterial strain (E. coli LE392 for λ) in 5 ml screw cap glass bottles. After allowing 20 min for adsorption, 3 ml of top agar (0.5% (w/v) agar) at 50°C was added and the contents poured onto a bottom, dry DDA agar plate. Plates were incubated without inversion at 37°C overnight. The following day two plaques were removed by taking a plug of agar using a Pasteur pipette. These plaques were suspended in 1 ml of TMG buffer and after adding 2 drops of $CHCl_3$ were vortexed using a whirly mixer for 1 min. The agar and cell debris was then pelleted (microfuge, high speed, 1 min) to leave supernatant lysate with a phage titre of approximately 10^5 pfu/ml.

Varying volumes (10, 50 and 100 µl) of this stock were added to 200 µl of sensitive bacteria (grown to an absorbance of A600 = 1.0) in LB containing $MgSO_4$ (10^{-2} M). After allowing 20 min adsorption, 3 ml of top agar (0.25% (w/v) agar) was added (at 50°C) and the contents poured onto a fresh, wet agar plate. Bacterial lawns were then incubated upright for 15 hr. For plates showing confluent lysis after overnight incubation at 37°C (as

compared with a phage free control), the top agar was harvested using a bent Pasteur pipette. The surface of the plate was washed with 2 ml of TMG. This liquid was then combined with the top agar and 500 µl of $CHCl_3$ to produce a slurry. This mixture was mixed vigorously by vortexing in a 25 ml universal tube for 10 to 15 min. The agar and cell debris was pelleted by centrifugation (MSE Multex 5 K, 4°C for 10 min) and the supernatant (lysate) decanted into a clean universal tube. Lysates were stored at 4°C over a few drops of $CHCl_3$. This method yielded lysates of up to 5×10^{11} pfu per ml for most λ derivatives.

For making Ecc bacteriophage lysates, LB based agar and incubation temperatures of 30°C were used throughout. For E. coli λ lysates DDA based agar and incubation temperatures of 37°C were used.

2.9. Screening for mutant extracellular enzyme phenotypes

Colonies resulting from mutagenesis procedures were inoculated onto NA plates (containing appropriate antibiotics) and incubated overnight at 30°C. These colonies were then transferred to extracellular enzyme detection plates (Pel, Cel and Prt), a MM plate and finally a NA plate using a replicating block. Enzyme detection plates were incubated at 25°C and 33°C. NA and MM plates were incubated at 30°C.

After 36 hr, enzyme detection plates were developed as described in section 2.10.1.

2.10. Quantitative enzyme assays

Samples used in the following assays were prepared as described in sections 2.16. and 2.17. For all enzyme assays, enzyme free controls were carried out in parallel. The control used in all the methods used was PM, the same medium in which the samples were grown.

2.10.1. β-galactosidase (β-gal) assay

The method was obtained from Hinton (pers. comm.) and has been described previously by Miller (1972).

The following were mixed in 1.5 ml microfuge tubes:

A) 250 µl supernatant
   250 µl Z buffer

B) 250 µl periplasm
   250 µl Z buffer

C) 50 µl sonicate
   250 µl H_2O + 250 µl Z buffer

To the reaction mixtures, 100 µl of ONPG (4 mg/ml freshly prepared in Z buffer) was added. Tubes were mixed and incubated at 37°C until a faint yellow colour appeared. At this stage the reaction was stopped by the addition of 250 µl of 1M Na_2CO_3 and the time recorded. The A_{420} was measured using H_2O as a reference. Units of activity were expressed as: A_{420} min/ml.

Z buffer (per litre)

8.52 g Na_2HPO_4 anhydrous

6.24 g $NaH_2PO_4.2H_2O$

0.75 g KCl

0.25 g $MgSO_4.7H_2O$

2.70 ml β-mercaptoethanol

[pH 7.0]

2.10.2. β-lactamase (Bla) assay

Bla was assayed as described by Hinton and Salmond (1987). The following reagents were mixed in (1 ml) microcuvettes:

A) 100 µl sonicate

700 µl 0.1 M phosphate buffer (pH 7.0)

B) 50 µl periplasm

750 µl 0.1 M phosphate buffer (pH 7.0)

C) 100 µl supernatant

700 µl 0.1 M phosphate buffer (pH 7.0)

To the reaction mixture, 20 µl of the chromogenic substrate (Nitrocephin 4 mg/ml in DMSO) was added. The microcuvettes were covered with a small piece of 'Parafilm' and mixed by inversion. The change in absorbance at 500 nm and 30°C was monitored using a Philips PU 8720 scanning spectrophotometer using the kinetics setting. Units of activity were expressed as: A_{500}/min/ml.

2.10.3. Pectate lyase (Pel) assay

This assay follows the breakdown of polygalacturonate (PGA) to digalacturonate (UDG) which leads to the increase in absorbance at 235 nm. This method was obtained from Hinton (pers. comm.) and was performed according to Chatterjee et al. (1985b).

The following reagents were added to a (400 µl) quartz cuvette:

292 µl reaction mix (pre-heated to 37°C)

7.5 µl supernatant/sonicate/periplasm

Cuvettes were mixed by inversion (using 'Nescofilm') and inserted into the temperature controlled cuvette holder of a Philips PU 8720 scanning spectrophotometer using the kinetics setting. The absorbance at 235 nm was followed. The initial reaction rate was obtained from the gradient of the slope during the early stages of the reaction. Units of activity were expressed as: A_{235}/min /ml.

Reaction mix	*Reaction buffer
3.45 ml reaction buffer*	76.8 ml H_2O
3.20 ml 0.57% PGA	23.0 ml Tris.Cl 1 M, pH 8.5
1.13 ml H_2O	78 µl $CaCl_2$

Note: 0.57% PGA was made by diluting a 2% stock solution (prepared for enzyme detection plates) with H_2O.

2.10.4. Protease (Prt) assay

This assay has been described by Ji et al. (1987). Protease activity was monitored by the release of the 'azo' dye from the substrate azocasein. Azocasein (2% w/v in H_2O) was stored in 10 ml aliquots at -20°C.

The following reagents were mixed in 5 ml glass test tubes:

A) 534 µl sonicate
   266 µl H_2O
   200 µl 1 M Tris.HCl pH 8.0

B) 667 µl supernatant
   133 µl H_2O
   200 µl 1 M Tris.HCl pH 8.0

C) 667 µl periplasm
   133 µl H_2O
   200 µl 1 M Tris.HCl pH 8.0

The reaction was initiated by the addition of substrate (1 ml of 2% (w/v) azocasein and incubating the tubes at 30°C. Samples (750 µl) were taken at t=0 hr and t=4 hr and added to microfuge tubes containing 375 µl of 14% (w/v) perchloric acid to stop the reaction. These tubes were then centrifuged (high speed, microfuge) for 3 min and the supernatant removed and transferred to microcuvettes containing 75 µl of 10 M NaOH. The contents were then mixed and the A_{436} nm recorded using H_2O as a reference. Activity of samples was expressed as change in $\Delta A436/h/ml$.

2.10.5. Polygalacturonase (Peh) assay

This assay was performed as described by Collmer et al. (1982). The following reagents were mixed together in 1.5 ml microfuge tubes:

450 µl Peh assay buffer

500 µl 1% PGA

50 µl cytoplasm/periplasm/supernatant

At t=0 min, 200 µl of the reaction mix was removed and transferred to a 10 ml test tube containing 200 µl of 'copper reagent' in order to stop the reaction. The remainder of the reaction mix was incubated at 30°C for 30 min. At t=30 min, a sample (200 µl) was removed from the reaction mix and transferred to a separate 10 ml test tube containing 200 µl of copper reagent. The samples containing copper reagent from the starting point (t=0 min) and the end point (t=30 min) were covered with glass marbles and boiled for 15 min. The tubes were then cooled before the addition of 200 µl of 'colour reagent'. The samples were then mixed by vortexing and diluted by the addition of 1.2 ml of H_2O. A 1 ml sample was removed and centrifuged for 2 min using a microfuge (high speed). Supernatants were decanted with care into 1 ml microcuvettes and their absorbance at 500 nm read using H_2O as a reference. Data was expressed as: A_{500}/min/ml.

Peh assay buffer

20 ml 4 M NaCl

20 ml 1 M Na acetate (pH 5.0)

1.6 ml 0.5 M EDTA

138.4 ml H_2O

Peh 'colour' reagent

Ammonium molybdate (25 g) was dissolved in H_2O (450 ml) for 30 min. Concentrated H_2SO_4 (21 ml) was then added. A solution of sodium arsenate was made separately by dissolving 3 g in 25 ml of H_2O. The two solutions were mixed and allowed to stand at 37°C for at least 48 hr after which time a yellow/green colour developed.

Peh 'copper' reagent

A solution containing 24 g of Na_2CO_3 and 12 g of potassium sodium tartrate in 250 ml H_2O was made. To this, 40 ml of 10% (w/v) $CuSO_4.5H_2O$ and 16 g of $NaHCO_3$ was added. A solution containing 18 g of Na_2SO_4 in 500 ml of hot H_2O was made separately and boiled for 30 min in order to expel air. After cooling, the two solutions were mixed 1:1 (v/v) and made up to 1 l with distilled H_2O.

2.10.6. Cellulase (Cel) assay

2.10.6.1. β-glucan cellulase assay plates

This method was obtained from D. Whitcombe (pers. comm.). Plates were made containing the following ingredients:

50 mM Tris.HCl pH 7.0

0.05% (w/v) Barley β-glucan 0.004% (w/v) Congo red

1.5% (w/v) agar

Plates were poured to a thickness of 3 mm and allowed to dry overnight at RT. Using an inverted Pasteur pipette and a suction pump, plugs

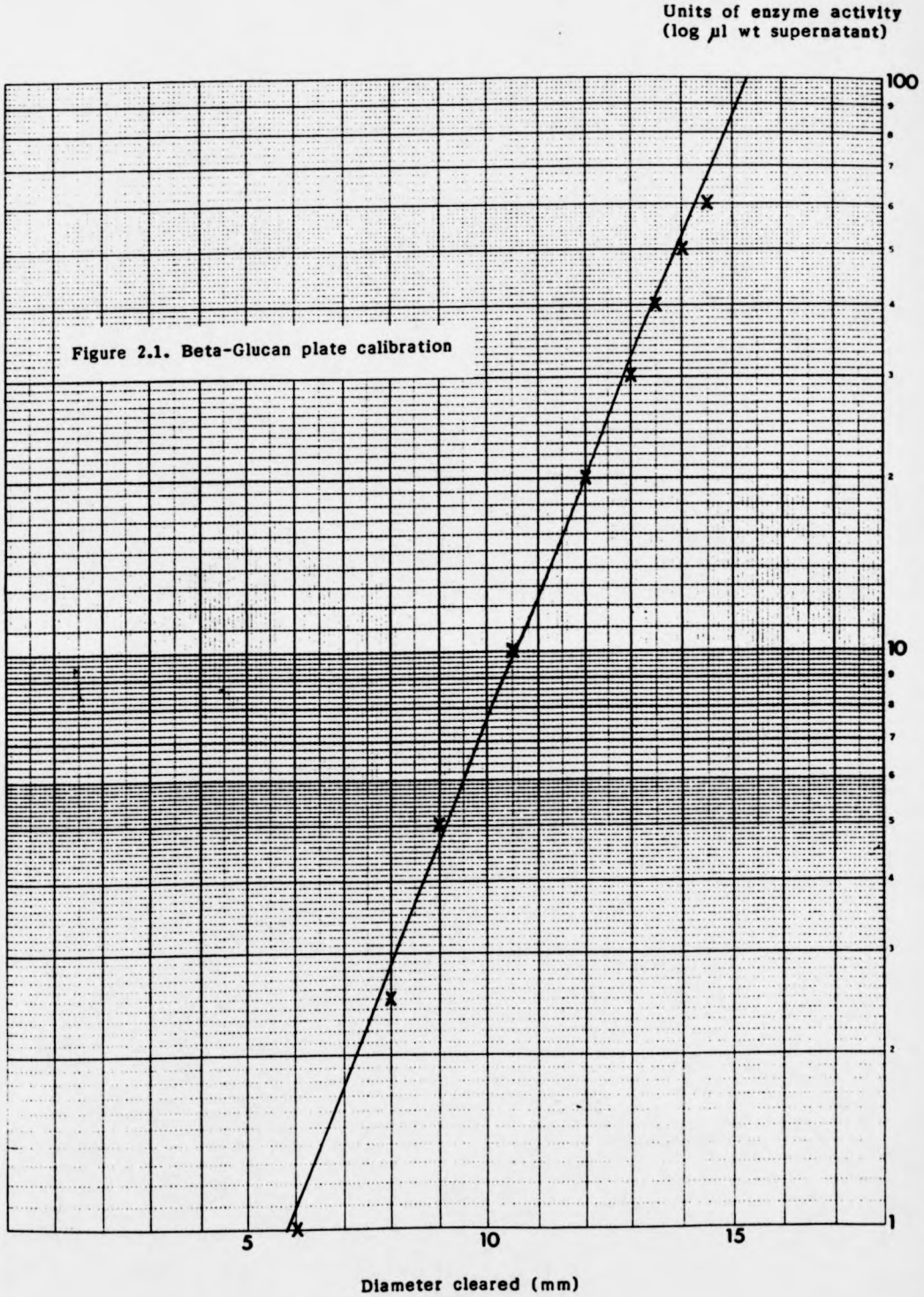


Figure 2.1. Beta-Glucan plate calibration

of agar were removed from the plates leaving loading wells. Samples of supernatant/cytoplasm/periplasm (10 µl) were placed into the wells and the plates incubated for 16 hr at 30°C. The diameter of the halo produced after this incubation was recorded. The diameter produced by the samples was compared to those produced by a dilution series of wild type Ecc supernatant. Data were expressed as mm/µl.

2.10.6.2. Carboxymethyl cellulose (CMC) spectrophotometric assay

Cel was assayed using a method obtained from Hinton (pers. comm.), adapted from Boyer et al. (1987).

The following reaction mix was set up:

0.33 ml CMC mix (1% w/v CMC in phosphate buffer (25 mM pH 7.0))

0.166 ml of sonicate/supernatant/periplasm sample.

At t=0 hr, 200 µl of the reaction mix was removed and transferred to a 10 ml test tube containing 500 µl of 'copper reagent' in order to stop the reaction. The remainder of the reaction mix was incubated at 30°C for 30 min. At t=4 hr, a sample (200 µl) was removed from the reaction mix and transferred to a separate 10 ml test tube containing 500 µl of 'copper reagent'. The samples containing copper reagent from the starting point (t=0 hr) and the end point (t=4 hr) were covered with glass marbles (in order to prevent evaporation) and boiled for 15 min. The tubes were then cooled before the addition of 1 ml of 'colour reagent'. The samples were then mixed by vortexing and diluted by the addition of 7 ml of H_2O. A 1 ml sample was removed and centrifuged for 2 min using a microfuge (high speed). Supernatants were decanted with care into 1 ml microcuvettes and

assayed for reducing sugars at 623 nm using H_2O as a reference. Data was expressed as change in A_{623}/min/ml.

2.10.6.3. Ostazin brilliant red (OBR-Cellulose) assay

This assay was developed by F. Ellard (pers. comm.) from Biely et al. (1985). This assay is based upon the release of the OBR dye from OBR-cellulose upon its digestion by cellulase. After precipitation of intact substrate, the amount of dye released into the supernatant was determined using spectrophotometric means.

The following reaction mix was set up:

140 µl OBR-cellulose (20.5 mg/ml)

140 µl H_2O

7 µl 50 x phosphate buffer (Table 2.2)

70 µl supernatant/cytoplasm/periplasm

Samples were incubated at 30°C for 2 hr and then stopped by the addition of 350 µl of ethanol/acetone (2:1 v/v). Samples were then allowed to stand at RT for 10 min and then centrifuged at 'high speed' for 2 min using a microfuge. The supernatant was decanted into a 1 ml microcuvette and its absorbance at 550 nm measured using H_2O as a blank. Units of activity were expressed as A_{550}/min/ml.

2.11. 'Pathogenicity' tests

Soft rot potato tuber assays were performed according to Hinton (1986), with modifications. Strains were grown at 30°C with shaking, overnight, in LB using appropriate antibiotic selection. Potato tubers were

washed and surfaced sterilised for 10 min in NaOCl (5% available chlorine). A 5 mm diameter metal tube was used to bore a shallow hole (2-3 mm) in the tuber. The bacterial culture (10 µl) was introduced into the hole using a yellow tip and pipette. The site of inoculation was sealed using a glass cover slip and paraffin wax (heated to 60°C). Tubers were then wrapped in triple layers of wet blue tissue paper and cling-film. Individual tubers were then placed into polythene bags and incubated at 25°C for 2-3 days. Following incubation, the tubers were sliced along the axis of injection and the diameter of rot measured. Non-inoculated tubers, tubers infected with Ecc HC131 and E. coli DH1 were used as controls.

2.12. Phage adsorption assay

Bacterial strains were incubated overnight with shaking (14-16 hr) in LB (10 ml) at the appropriate temperature. Bacteriophages (10^9 pfu/ml) were added and allowed to adsorb for 20 min at RT. Bacteria were removed by centrifugation (5 K for 10 min) and the supernatant treated with $CHCl_3$. The number of unadsorbed bacteriophage remaining in the culture supernatant (post adsorption) was determined by titrating the supernatant on a lawn of sensitive bacteria.

2.13. Cosmid technology

2.13.1. Packaging cosmids into bacteriophages in vivo

This method describes the 'amplification' of a cosmid library which had previously been introduced into E. coli DH1. The procedure was modified from a method described by White et al. (1983).

A 10 ml culture of E. coli DH1 containing an heterogeneous Ecc

library was grown overnight at 30°C with shaking. A 250 ml flask containing 25 ml of LBAp ($MgSO_4$ 10^{-2} M) was inoculated with this culture (250 µl) and incubated at 30°C with shaking (260 rpm). When an A600 of 0.3 was reached, 250 µl of λ cI857 (10^{10} pfu/ml) was added. The culture was then left static for 30 min at 30°C. The flask was then shifted to 42°C and left static for a further 30 min. The flask was then shaken (260 rpm) at 37°C until lysis occurred, approximately 3 hr. At this stage 1 ml of $CHCl_3$ was added and the culture shaken for a further 5 min. The contents of the flask were then transferred into a 20 ml universal tube and centrifuged at 5 K for 10 min at 4°C. The supernatant was then decanted into a universal tube and stored over $CHCl_3$ at 4°C until needed.

2.13.2. Transduction using λ cosmid lysates

Cosmids and transposons were introduced into LamB+ Ecc strains using λ as a delivery vehicle. Ecc cells were grown overnight in LB containing appropriate antibiotics (to maintain lamB carrying plasmids) and $MgSO_4$ (necessary for λ adsorption) at 30°C. Cells were harvested by centrifugation (5 K, 5 min) and then resuspended in 500 µl of the λ cosmid lysate. After 20 min static adsorption at RT, a further 10 ml of LB was added. Infected cells were then transferred to a shaking incubator (30°C, 275 rpm) and incubated for 1 hr to allow expression of cosmid encoded antibiotic resistance genes. After expression, cells were harvested by centrifugation (5 K, 5 min), resuspended in 1 ml of LB and spread onto NA plates (100 µl aliquots) containing appropriate antibiotics to select for cosmid transductants. It was necessary to incubate plates for 2 days at 30°C before colonies appeared.

2.14. Conjugal transfer of plasmids by patch mating

Newly formed colonies of donor and recipient strains were mixed on a dry NA plate using a sterile wire loop. Plates were then incubated at 30°C for 24 hr. After incubation, bacteria from the patch mating area were streaked out onto appropriate plates to select the recipient strain and counterselect the donor. Ecc plasmid recipients were separated from auxotrophic donors using a plasmid encoded resistance marker. In order to counterselect against auxotrophic E. coli donors, transconjugants were plated onto MM plates containing sucrose as the carbon source.

2.15. Curing plasmids from Ecc

Ecc strains containing pBR322 based plasmids were propagated in LB for 2 days without antibiotics necessary for plasmid maintenance. A fresh culture was then inoculated (1:100 v/v) using this bacterial culture and incubated for a further 2 days. A loopful of culture was then streaked out onto NA plates to produce single colonies. The resulting colonies (50-200) were then tested on NA and NA containing the appropriate antibiotic resistance to identify cured colonies.

2.16. Fractionation of cultures into supernatants and whole cell sonicates

Cells were inoculated into 5 ml of Pel inducing medium (PM) in 25 ml universal tubes, placed horizontally into an orbital shaker (30°C, 200 rpm) and incubated overnight. The following day the O.D.600 of the culture was recorded. The cells were centrifuged at 5 K for 5 min at 4°C using a MSE Multex centrifuge. The supernatant was decanted off and stored at 4°C. The pellet was washed once in the same medium and then re-pelleted as before. The pellet was then resuspended in 5 ml of the same medium. This cell

suspension (5 ml) was then sonicated in order to release the intracellular contents. Sonication was carried out on ice using 25 ml beakers and a large (3 cm) probe. Cells were subjected to 3 x 30 sec cycles of sonication at an amplitude of 6 microns (peak to peak) with 30 sec cooling down intervals between cycles. After sonication, the samples were centrifuged (10 min, 4°C 5 K, MSE Multex) in order to remove cell debris and unlysed cells. All samples (sonicate and supernatant fractions) were frozen in aliquots (1 ml) at -20°C and defrosted slowly and fully before use.

2.17. Fractionation of cell cultures into cytoplasms, periplasms and supernatants

Two different methods for cellular fractionation will be described. The most reliable and reproducible method was by generating spheroplasts. A third method was also attempted which employed the use of polymixin B; however, Ecc was resistant to this compound.

2.17.1. Osmotic cold shock

The method used was from Neu and Heppel (1965). The A_{600} of overnight cultures (5 ml) grown in PM (30°C with shaking) was recorded. After the addition of Tris.Cl 0.5 M, pH 7.8 the culture was incubated for 10 min at RT. The culture was then centrifuged for 10 min at 5 K, 4°C and the supernatant collected and stored at 4°C. The cell pellet was then resuspended in 800 µl of solution 1 and transferred to a 1.5 ml microfuge tube. The cell suspension was incubated at 30°C for 10 min and then pelleted by centrifugation at 12 K for 1 min using a microfuge. The supernatant was carefully removed and the pellet immediately and rapidly resuspended in 500 µl of ice-cold H_2O. The sample was subjected to 10 min incubation on

ice and then pelleted at 12 K for 3 min, after which time the supernatant was removed (periplasmic fraction) and stored at 4°C. The cell pellet was then resuspended in Tris.Cl 50 mM, pH 7.8 and sonicated to produce the cytoplasmic fraction. For long term storage, samples were frozen at -20°C in 1 ml aliquots. After storage, samples were fully but slowly de-frosted on ice and never re-frozen.

2.17.2. Fractionation of Ecc by generating spheroplasts

This method releases the soluble periplasmic proteins into the supernatant by stripping off the OM without rupturing the IM. A method designed to prepare *E. coli* spheroplasts (Osborn and Munson, 1974) was adapted for *Ecc* as part of this work.

Bacterial cultures were grown by inoculating a fresh colony into a 25 ml universal tube containing PM (5 ml), LB (25 µl) and appropriate antibiotics. Cultures were incubated overnight at 30°C with shaking (200 rpm). The following day the A_{600} of the culture was recorded and the cells centrifuged in order to obtain the 'supernatant' fraction. The pellet was then resuspended in 5 ml of TS solution and the cell suspension placed on ice. Immediately, 50 µl of lysozyme solution was added followed by incubation on ice for 2 min. The cell suspension was then transferred to a 50 ml beaker and gently stirred at 4°C using a magnetic stirrer. A solution of EDTA was added very slowly but continuously (over 10 min) to the cells using a 10 ml pipette with a a rubber tube attached at the nozzle. The rate of flow from the nozzle was controlled using an adjustable clip. After addition of the EDTA solution was complete, a 5 µl aliquot of cells was placed onto a slide and observed at high magnification using 'phase contrast' light microscopy. It was then possible to determine the extent of

spheroplasting (and therefore periplasmic release). If < 70% spheroplasting was observed the cells were incubated at 37°C for 10 min, after which time > 90% of the cell population usually existed as spheroplasts. The spheroplasts were then centrifuged at 5 K, 4°C for 10 min using a MSE Multex centrifuge. After centrifugation the cell pellet was always associated with a 'stringy' thread of cell debris which indicated that spheroplasting had been successful. A sample of supernatant (5 ml) was removed taking care not to disturb any of the spheroplasts or debris. This sample was the 'periplasmic' fraction. The remainder of the supernatant was decanted off and discarded. The pellet was resuspended in PM (5 ml). This sample was sonicated and centrifuged to obtain the 'cytoplasmic' fraction (see section 2.16.).

TS solution

Tris.HCl 10 mM, pH 7.8

Sucrose 0.75 M

EDTA solution

2 ml 15 mM EDTA pH 8,0

8 ml H_2O

Lysozyme solution

2 mg/ml lysozyme in Tris.HCl 10 mM,pH 7.8

PM

Pel minimal medium (see Table 2.4.)

2.18. Isolation of plasmid DNA

2.18.1. Isolation of plasmid DNA from Ecc

Ecc plasmid DNA was isolated using a method developed for preparing plasmid DNA from Streptomyces sp. (D. Hood, pers. comm.), adapted from Kieser (1984). This method was used to produce a cleared cell lysate which was then used to transform E. coli. Plasmid DNA was then prepared from E. coli (section 2.8.3.1.) and then analysed.

An overnight culture of cells (5 ml) was pelleted by centrifugation (5 K, 5 min) and the pellet washed once with TMG buffer (section 2.8.). The pellet was then resuspended (using a 2 ml microfuge tube) in 1 ml of freshly prepared lysis mix (0.3 M sucrose, 25 mM Tris (pH 8.0), 25 mM EDTA (pH 8.0), lysozyme (2 mg/ml), and incubated on ice for 20 min before adding 500 µl of SDS solution (0.3 M NaOH, 2% (w/v) SDS). The cells were then agitated using a vortex until lysis was evident (usually instantaneous). The top of the tube was pierced with a hot needle and the tube incubated at 65°C for 10 min. The sample was then cooled on ice and extracted with 160 µl of acid phenol/$CHCl_3$ (made as in section 2.23.1. but not equilibrated). The aqueous phase (upper) was then mixed with 140 µl of 3 M unbuffered sodium acetate and 1.4 ml of isopropanol (IPA). The DNA was allowed to precipitate for 10 min on ice and was then pelleted at 12 K for 10 min using a microfuge. The DNA pellet was washed with 70% (v/v) EtOH, dried by desiccation and resuspended in 50 µl TE buffer. This DNA solution was used to transform E. coli strains (section 2.6.).

2.18.2. Rapid small scale isolation of plasmid DNA from E. coli

2.18.2.1. The Boiling method

This method, described by Maniatis et al. (1982), is fast and convenient. It is especially useful when analysing large numbers E. coli strains carrying recombinant plasmids.

E. coli transformants were inoculated into 2 ml of LB (plus appropriate antibiotics) in 10 ml test tubes and incubated at 37°C with good aeration, using a New Brunswick Scientific G24 shaking incubator. The culture (1.5 ml) was transferred to a 1.5 ml microfuge tube and centrifuged for 2 min at high speed using a microfuge. The culture supernatant was then carefully removed by aspiration and the pellet resuspended in 0.35 ml of Triton solution (8% sucrose, 0.5% (v/v) Triton X-100, 50 mM EDTA pH 8.0, 10 mM Tris.HCl pH 8.0) before adding 0.25 ml of a freshly prepared solution of lysozyme (10 mg/ml in 10 mM Tris.HCl pH 8.0). The tube was gently mixed by inversion, incubated at RT for 5 min and then placed in a boiling water bath for 1 min in an microfuge boiling rack. The tube was then centrifuged for 20 min at high speed in a microfuge to produce a firm gelatinous pellet which was removed using a toothpick. An equal volume of isopropanol (IPA) (0.2 ml) was added to the supernatant and after mixing the solution was stored for 20 min at -20°C or until 'cloudy' in appearance. The solution was centrifuged for 20 min at high speed (microfuge) and the supernatant decanted off. The tube was then centrifuged for a further 30 sec and the remaining supernatant removed using a yellow tip. The pellet was washed once with 0.2 ml 70% (v/v) ethanol and then dried under vacuum for 10 min. The pellet was resuspended in 100 µl TE buffer and stored at -20°C. For restriction analysis 5 µl of this DNA solution was used, and for

transformation 2 µl was used. For long term storage at -20°C the supernatant was extracted once with phenol/$CHCl_3$ prior to IPA precipitation.

2.18.2.2. Small scale (mini-prep) plasmid DNA preparation by alkaline lysis

The method used has been described by Maniatis et al. (1982) and was used for the rapid extraction of plasmid DNA from small (5 ml) or medium sized (50-100 ml) culture samples.

E. coli cells were grown overnight at 37°C, with shaking, in 5 ml of LB containing appropriate antibiotics. Bacterial cultures were routinely grown in horizontally positioned 25 ml screw cap universal tubes. Cells were pelleted by centrifugation for 5 min at 5 K at RT, and resuspended in 2 ml of solution 1 (25 mM Tris.HCl pH 8, 50 mM glucose, 10 mM EDTA) containing lysozyme 2 mg/ml added just prior to use. This cell suspension was incubated on ice for 10 min before the addition of 0.4 ml of freshly prepared alkaline SDS solution (0.2 M NaOH, 1% (w/v) SDS) [alkaline SDS was prepared from stocks of SDS (20% w/v) and NaOH 10 M]. After rapid inversion 3 or 4 times on ice, the contents of the tube became clear and viscous due to cell lysis. Ice cold sodium acetate (0.3 M/5M; made by mixing 60 ml 5 M sodium acetate, 28.5 ml H2O and 11.5 ml glacial acetic acid) was added, and the mixture was gently mixed by inversion and incubated on ice for 5 min. The tube was then centrifuged to remove unlysed cells, membrane-bound chromosomal DNA and other cell debris. Then, the supernatant was transferred to a fresh tube containing an equal volume of IPA, mixed by vortexing and centrifuged for 2 min at high speed using a microfuge. The upper aqueous phase was removed with care (so as to avoid the white precipitate at the interface) and transferred to a new microfuge tube containing the same volume of chloroform: isoamyl alcohol (24:1). After

mixing by vortexing and separating the two phases by centrifugation for 2 min at high speed using a microfuge, the upper aqueous phase was removed and transferred to a fresh tube. Ammonium acetate (7.5 M) 1/10 volume and IPA 4/5 volume were added and the solution was mixed by inversion and allowed to stand for 20 min at -20°C. The tube was then centrifuged at high speed for 20 min and the supernatant removed by decanting, re-spinning tubes and removing residual supernatant with a yellow tip and Gilson pipette. The pellet was washed with 1 ml of 70% (v/v) ethanol which was subsequently removed, as above, before drying under a vacuum for 10 to 20 min. After desiccation the pellet was resuspended in 50 µl of TE buffer and stored at 4°C or -20°C. To remove RNA, DNase free RNase (20 µg/ml) was added prior to the phenol extraction step. RNase treatment was carried out at 37°C for 30 min. In order to obtain higher purity DNA, phenol extractions were repeated until no white precipitate was present at the interface between the two phases, before continuing with the rest of the procedure.

2.18.3. Medium sized preparations (midi-prep) of plasmid DNA

For 'midi-preparations' of plasmid DNA, 50 ml cultures of cells (at stationary phase) were used. These had been grown in 500 ml flasks at 37°C in an orbital shaker (275 rpm). Cells were pelleted using a MSE High Spin-21 centrifuge and the supernatant was removed and discarded. The cell pellet was resuspended in 2 ml lysis solution (solution 1 in section 2.18.2.2.), transferred to an Oakridge tube (40 ml screw cap plastic centrifuge tube), and incubated for 10 min on ice.

The rest of the procedure was exactly the same as for the mini-prep (alkaline lysis) protocol but was scaled up in proportion to the cell input

volume. The DNA pellet was finally dissolved in 200-400 µl of TE buffer and stored at 4°C or -20°C. Centrifugations were carried out using 8 x 50 rotor and a MSE High Spin-21 centrifuge.

2.18.4. Large scale (maxi-prep) plasmid preparations

This method was used to prepare high yields of high quality plasmid DNA. It was also used to prepare RNA free DNA without the use of RNase. The protocol was adapted from that of Maniatis et al., 1982 (S. McGowan, pers. comm.).

Cells were grown in 200 ml of 2YT containing appropriate antibiotics in 2 l flasks at 37°C. Cells were either grown overnight under these conditions or, when appropriate, plasmids were amplified by the addition of spectinomycin (300 µg/m) at an absorbance of A600 = 0.6 to 0.8 before incubation for a further 18 hr. Cells were harvested using an MSE High Spin-21 centrifuge using a 6 x 300 rotor, for 10 min at 10 K and 4°C. The pellet was washed in 20 ml TES (50 mM Tris.Cl pH 8.0, 5 mM EDTA pH 8.0, 50 mM NaCl), re-centrifuged and stored at -20°C for 30 min or until required. The pellet was then resuspended in 10 ml of STE (25% (w/v) sucrose, 50 mM Tris.Cl pH 8.0, 5 mM EDTA pH 8.0) by vigorous vortexing. Fresh lysozyme solution (1 ml) (10 mg/ml lysozyme, 0.25 M Tris.HCl pH 8.0) was added before incubation at 4°C for 5 min with occasional mixing by rocking. At this stage lysis was often evident as seen by the solution becoming viscous. A 2.5 ml solution of EDTA (0.5 M, pH 8.0) was added and the cell suspension was left for a further 5 to 10 min. Using a 25 ml pipette and pipette-pump, 16 ml of Triton lysis mix (0.1% (v/v) Triton X-100, 50 mM Tris.HCl pH 8.0, 5 mM EDTA pH 8.0) was added rapidly and the solution rapidly mixed by shaking vigorously. After a further 20 min

incubation on ice, the lysed cells were centrifuged for 30 min at 19 K 4°C in an MSE High Spin-21 centrifuge. The supernatant was decanted through muslin into a 200 ml measuring cylinder containing 28.5 g of CsCl. The volume was brought up to 38 ml with TES and the CsCl dissolved completely by rocking, after the tube had been sealed with parafilm, and occasional 1 min incubations at 60°C. The resulting solution was transferred to an Oakridge tube before adding 2 ml of ethidium bromide (EtBr) (5 mg/ml) and then allowed to stand on ice for 10 min. The tube was then centrifuged (20 min, 15 K, 4°C) and the supernatant decanted (through a 20 ml syringe containing siliconised glass wool and 16 gauge needle) into a Beckman heat-seal Vti 50 centrifuge tube. The tube was then topped up to the neck using CsCl solution and heat sealed. Balanced tubes were centrifuged using a Vti 50 rotor at 45 K, 22°C for 14 hr. The plasmid band (usually the lower of two bands) was identified using long wave UV visualization and removed by first piercing the top of the tube with a 16 gauge syringe needle and then carefully removing the plasmid band using the same gauge needle and a 10 ml syringe. The plasmid band (5 ml) was then transferred to a Vti 65 tube (5 ml) and centrifuged in a Vti 65 rotor, 65 K, 22°C for 5 to 6 hr. This step, although not essential, was used to further purify the plasmid band. The plasmid band was removed as before except using a 21 gauge syringe and 1 ml syringe collecting 0.5 to 1 ml of sample. The EtBr was removed by extracting the sample 5 times with an equal volume of salt-saturated IPA. Plasmid DNA was precipitated by the addition of 2 volumes of H_2O and 6 volumes of ethanol. (Note: the sample was diluted with H_2O so as not to precipitate the CsCl.) The ethanol precipitation was carried out at -20°C for 1 to 24 hr after which the sample was centrifuged for 30 min at 15 K, 4°C in an MSE High Spin-21 centrifuge. The supernatant was removed and the

location of the transparent pellet marked on the tube using a marker pen. The pellet was washed with 70% (v/v) ethanol in order to remove any residual CsCl and re-centrifuged for 10 min at 20 K, 4°C. After decanting the supernatant, the pellet was dried under vacuum for 30 min, dissolved in 0.5 ml of TE buffer and stored at 4°C or -20°C.

2.19. Restriction endonuclease digestion of plasmid DNA

DNA was digested with the appropriate endonucleases using buffers and conditions as recommended by the suppliers (BRL, NEB, Bohringer Mannheim and Amersham International). Reactions were generally carried out in 20 µl total volumes and DNA fragments then analysed by agarose gel electrophoresis or used as required. Where digestion using more than one enzyme was required, and if the digestions were incompatible, two-step reactions were carried out. The first step was carried out at the lower salt concentration or temperature and before initiating the second step the salt concentration was altered accordingly with 1 M NaCl and/or the temperature raised. All reactions were carried out for at least 60 min.

2.20. Agarose gel electrophoresis

Horizontal agarose gels were prepared by melting agarose in TAE electrophoresis buffer (0.04 M Tris-acetate, 0.001 M EDTA). After the addition of EtBr (0.5 µg/ml final conc.), the agarose solution was cooled to 60°C and poured into gel moulds (Biorad Sub Cells or mini Sub Cells). Gels (when set) were placed in the appropriate gel tank and submerged in TAE containing EtBr (0.5 µg/ml final conc.). DNA samples were prepared for electrophoresis by adding 0.1 volume of loading buffer (0.25% (w/v) bromophenolblue (BPB), 50% (v/v) glycerol) and loaded with care into gel

slots. Gels were electrophoresed (80 V mini Sub Cells or 100 V Sub Cells) until adequate separation of DNA bands had occured. For high quality resolution gels were run at 20-30 V for 15 hr). After electrophoresis, gels were observed under long wave UV and photographed using a UV transilluminator and Polaroid positive/negative 665 film.

2.21. Preparation of DNA fragments from agarose gels

The fragment of interest was identified by observation under long wave UV light. A trough extending the length of the DNA band and approximately 2 mm thick was cut into the agarose gel to the base plate using a scalpel. After lowering the level of buffer to just below the level of the gel, the DNA band was electrophoresed into the trough using 60 sec cycles of 100 mV. After each cycle the buffer from the trough was collected and replaced with fresh buffer. The migration of the band was monitored until it had all been collected. The DNA was phenol extracted (section 2.23.2.) and precipitated using IPA (section 2.24.2.).

2.22. De-phosphatasing linear double stranded DNA

De-phosphatasing was carried out to prevent re-circularisation of vector DNA during ligation reactions. After digestion of vector DNA the volume was increased from 20 µl to 50 µl using distilled H_2O and 5 µl of de-phosphatase buffer (from Bohringer Mannheim). Calf intestinal alkaline phosphatase (1 µ) was added and the reaction allowed to proceed for 30 min at 37°C. The sample was then extracted twice with phenol/$CHCl_3$ (section 2.23.). DNA was precipitated by the addition of 1 ml of ethanol and incubation on ice for 30 min. The DNA was pelleted by centrifugation (high

speed, microfuge, 20 min), dried under vacuum (10-20 min) and resuspended in 10 µl of TE buffer.

2.23. Extraction of protein from DNA solutions using neutral phenol/chloroform

2.23.1. Preparation and storage of neutral phenol/chloroform

Phenol/chloroform ($CHCl_3$) mix was prepared (with care) by dissolving 100 g of phenol crystals and 100 mg of 8-hydroxyquinoline in 100 ml of $CHCl_3$)/IAA (24:1 (v/v)). The phenol/$CHCl_3$/IAA solution was neutralised by shaking with 3 changes of Tris.HCl 1 M, pH 8.0 or until the aqueous layer was pH 8.0. The neutral phenol/$CHCl_3$ mix was stored under TE buffer at 4°C in the dark.

2.23.2. Use of phenol/chloroform

DNA solutions were mixed with the same volume of phenol/$CHCl_3$ (prepared as in section 2.23.1.) and mixed until an emulsion formed. This was achieved by vortexing small samples (1-10 ml) or repeatedly inverting large samples (> 10 ml) The two phases (organic and aqueous) were re-formed by centrifugation. The upper (aqueous) phase was removed with care so as not to disturb the interface which usually contained white particulate matter. This process was repeated until the interface between the two phases was clear. The aqueous phase was then extracted with the same volume of $CHCl_3$/IAA (24:1) using the same method as above. The DNA was recovered by ethanol or IPA precipitation (section 2.24).

2.24. Precipitation of plasmid DNA

2.24.1. Precipitation of plasmid DNA using ethanol (EtOH)

To a DNA solution, half a volume of cold ammonium acetate (7.5 M, pH 7.5) and 3 volumes of cold EtOH were added. The resulting solution was shaken (or vortexed for small (<10 ml) volumes) until thoroughly mixed and incubated at -20°C for 1 to 24 hr. DNA was recovered by centrifugation at high speed (10 K) for 20 min. The DNA pellet was washed with 1-5 ml of 70% (w/v) EtOH. The sample was centrifuged for 10 min at high speed in order to reform the DNA pellet. The supernatant was removed and the pellet dried under vacuum. The DNA pellet was resuspended to the desired concentration in TE buffer and stored at either 4°C or -20°C.

When precipitating very small volumes of DNA (e.g. after phenol extracting a 20 µl restriction digest reaction), 1 ml of EtOH was used without the addition of salt (which is present in the restriction enzyme buffer). The remainder of the procedure was as above.

2.24.2. Precipitation of plasmid DNA using isopropanol (IPA)

IPA precipitation was used when it was more convenient to work with a smaller sample volume. This is because DNA precipitation can be achieved with 0.7 volumes of IPA whereas 2 volumes of EtOH are required.

The DNA solution was mixed with 0.2 volumes of ammonium acetate (7.5 M) and 0.8 volumes of IPA. The solution was mixed well and incubated at -20°C for approximately 1 hr. The DNA was then pelleted by centrifugation, washed with 70% (w/v) EtOH and dried under vacuum. The pellet was then resuspended to the desired concentration in TE buffer.

2.25. Ligation reactions

2.25.1. Re-circularisation of digested plasmid DNA

For re-circularisation, plasmid DNA was diluted after digestion with TE buffer to give a final DNA concentration of < 10 µg/ml. To remove residual restriction enzyme, the sample was extracted with an equal volume of phenol/$CHCl_3$ (section 2.23.2.) and precipitated with 1 ml of ethanol (section 2.24.1.). The DNA was dissolved in 20 µl TE buffer and stored at 4°C. Aliquots of DNA (3 µl) were heated to 65°C for 10 min and allowed to anneal slowly at room temperature for 10 min before storing on ice for 10 min. The DNA was then mixed with 1 µl 10 x ligation buffer (4 mM ATP, 60 MM MgCl2, 0.1 M DDT, 0.66 M Tris.Cl pH 7.6) and 5 µl of H_2O. Ligation was initiated by the addition of 1 µl (1 u/µl) of T4 ligase and carried out at 15°C for at least 14 hr.

2.25.2. Cloning into plasmid vectors

A fragment to vector concentration ratio of 4:1 was used in a reaction mixture volume of 10 µl. Vectors were de-phosphatased using calf intestinal phosphatase (section 2.22.). The vector and fragment were mixed and made up to 8 µl with H_2O and then heated to 65°C for 10 min. Samples were then cooled on ice for 5 min. To initiate the ligation reaction, 1 µl of T4 DNA ligase and 1 µl of 10 x ligation buffer (section 2.25.1.) were added. For 'sticky-end' ligations the reaction was incubated for 4 hr at RT. For a higher efficiency of ligation, reactions were carried out at 15°C for 16 hr. All blunt-ended ligations were carried out at 15°C for 16 hr.

2.26. Transformation of E. coli

E. coli was transformed using the standard method as described by Maniatis (Maniatis et al., 1982). Grade 1 $CaCl_2$ was used throughout. Bacterial strains were grown in rich medium such as 2YT.

2.26.1. Preparation and storage of competent E. coli cells using $CaCl_2$

E. coli strains to be transformed were incubated overnight at 37°C with shaking and subcultured the following day 1:100 into 1 l flasks containing 100 ml of the same media. Flasks were incubated at 37°C with good aeration until reaching A600 = 0.5-0.7. Cells were then centrifuged at 5 K for 5 min at 4°C and the pellet resuspended in half the original volume of $CaCl_2$ (0.1 M). Cells were then stored on ice for 10 min before pelleting and resuspended in 1/10 the original volume of $CaCl_2$ (0.1 M). After a further 10 min on ice the cells were re-pelleted and resuspended in 1/16 the original volume of $CaCl_2$ and stored on ice for at least 1 hr before use. For long term storage, cells were treated with $CaCl_2$ for approximately 5 hr, mixed with an equal volume of FCM (0.1 M $CaCl_2$, 50% (v/v) glycerol) and stored at -70°C in 500 µl aliquots.

2.26.2. Transformation protocol

To initiate transformation plasmid DNA (50-100 ng) was mixed with 0.1 ml competent E. coli cells (section 2.26.1.). The cell/DNA mixture was incubated on ice for 30 min and then heat-shocked by incubation at 42°C for 2 min. Fresh LB (1 ml) was added to the heat shocked cells which were then incubated with shaking at 37°C for 30-60 min to allow antibiotic expression. The transformed cells were pelleted by centrifugation (microfuge, high speed, 1 min). The supernatant was decanted to leave a residual volume of 0.1 ml

in which the pellet was resuspended. The transformed cells were then spread onto selective plates (to select transformants) and incubated overnight at 37°C.

2.27. M13 Dideoxynucleotide termination sequencing using randomly generated clones

2.27.1. Introduction

Random DNA sequencing was performed as described by Bankier et al. (1986). The target DNA molecule was first broken down into random fragments which were then cloned into M13. Single stranded M13 clones were then sequenced using the chain-termination procedure described by Sanger et al. (1977).

2.27.2. Isolation of target DNA

To release the target DNA (5 µg), 50 µl of recombinant plasmid prepared by the 'midi-prep' procedure (section 2.18.3.) was digested with the appropriate restriction enzyme. After agarose gel electrophoresis the band of interest was excised using a scalpel. The DNA was released using a Biorad electroelutor. Electroelution was carried out according to the manufacturers instructions. In this experiment the target DNA (3.7 Kb) was electroeluted at 100 V for 20 min. The DNA was precipitated using IPA and resuspended in a final volume of 20 µl TE buffer.

2.27.3. Circularisation of target DNA

This procedure was performed in order to ensure that the fragments produced in the sonication step (section 2.27.4.) were fully representative of the target DNA.

The fragment was self ligated using the entire DNA sample obtained using the method described in section 2.27.2. The DNA (20 µl) was mixed with 3 µl T4 DNA ligase (1 µ/ul) and 3 µl ligation buffer (section 2.25.1.)

and made up to a final volume of 30 µl with H_2O. The ligation reaction was carried out at RT (23°C) and allowed to proceed for 4 hr.

2.27.4. Sonication

Sonication was carried out using a cup horn sonicator. The output was set at maximum (10) on a continuous cycle. Sonication of the entire sample (30 µl) was carried out at 4°C for 3 x 60 sec cycles with 30 sec cooling intervals between cycles. After sonication the extent of DNA fracture was monitored by running 2 µl of the DNA sample in an agarose gel and comparing the size distribution with molecular weight markers (ØX174 digested with HaeIII).

2.27.5. End-repairing sonicated DNA

End-repair is necessary for cloning the damaged DNA fragments resulting from sonication. End-repair was carried out by incubating the sonicated DNA (30 µl) with DNA polymerase (Klenow fragment) (10 u), T4 DNA polymerase (10 u) and 2 µl 1.5 M DNTPs in a reaction volume of 40 µl. Incubation was carried out at RT for 30 min. MgCl2 (high grade) was also added to the reaction mix to give a final concentration of 5 mM.

2.27.6. Size selection

The sonicated/end-repaired DNA was electrophoresed on a 1% agarose gel using bacteriophage ØX174 cut with HaeIII as molecular weight markers. A gel slice covering the 300 to 600 bp size range was removed and the DNA recovered using a Biorad electroelutor. Care was taken to avoid contaminating the sample with the molecular weight markers which were blunt ended. The DNA was electroeluted at 100 V for 20 min until the gel

slice lost the DNA, as seen by monitoring with long wave UV light. The ammonium acetate solution containing the DNA was collected. DNA was then precipitated with an equal volume of IPA overnight at -20°C.

2.27.7. Preparation of 'high efficiency' competent cells

E. coli TG1 cells were made competent using the protocol of Hanahan (1983). This method gives an increased transformation frequency compared with $CaCl_2$ prepared cells. This is an important factor when using end-repaired/sonicated fragments in ligation reactions.

A colony of E. coli TG1, stored on a MMA plate at 4°C, was inoculated into 5 ml of 2YT broth and incubated overnight at 37°C with shaking. The following day, fresh 2YT broth (30 ml in a 250 ml flask) was inoculated with 0.3 ml of the TG1 culture and incubated with shaking (200 rpm) at 37°C until reaching an absorbance of A600 of 0.6 (approximately 2 hr). The culture was transferred to an Oakridge tube (40 ml screw cap plastic centrifuge tube) and centrifuged for 10 min at 3 K and 4°C using an MSE High Spin-21 centrifuge. After decanting the supernatant, cells were resuspended gently in 2.5 ml of cold Solution 1. The cells were incubated on ice for 15 min before adding DMSO (100 µl) and incubated on ice for a further 5 min. Solution 2 (100 µl) was then added followed by another 15 min incubation on ice. Finally a further aliquot of DMSO (100 µl) was added and after 5 min incubation on ice the cells were ready to use. The competent cells were used within 1 hr of preparation.

Solution 1 (TFB)

10 mM K-MES (pH 6.0)

100 mM RbCl

45 mM $MnCl_2.4H_2O$

10 mM $CaCl_2.2H_2O$

3 mM Hexamine cobalt III Chloride ($HACoCl_3$)

Solution 2

2.25 M DDT

40 mM potassium acetate pH 6.0

2.27.8. Cloning into M13 mp8

The M13 bacteriophage derivative, M13 mp8, was used as the cloning vector. Commercially prepared SmaI cut, de-phosphatased vector was purchased from Amersham International. This can be used for cloning blunt ended DNA fragments such as those produced by end-repair.

2.27.9. Ligation

The M13 mp8 vector (120 ng) was mixed with 35 µl H_2O and 12 µl of ligase buffer (from Bethesda Research Laboratories). The vector/buffer mix was then mixed in tubes 1 to 6 according to Table 2.6. Ligation was carried out at 15°C overnight (16 hr). After ligation the samples were diluted with 5 volumes TE buffer and stored at -20°C.

2.27.10. Transformation

Competent cells (200 µl) were mixed with 20 µl of diluted ligation mixture and incubated on ice for 45 min. The competent cell/ligation mix

Table 2.6. Ligation into M13 mp8

		Tube number					
		1	2	3	4	5	6
Vector/buffer[1]	8µl	8µl	8µl	8µl	8µl	8µl	8µl
DNA		1µl	2µl	3µl	-	-	-
Alu1 cut lambda (10ng/µl)		-	-	-	1µl	-	-
T4 DNA ligase[2]	1µl	1µl	1µl	1µl	1µl	-	-

1 Vector buffer mix was made by mixing;

6µl (10x) ligation buffer

6µl (120ng) *Sma1* cut M13

41µl water

2 1µl = 100 units of activity

Ligated DNA was introduced into *E. coli* TG1 by transformation

was then heat-shocked at 42°C for 3 min and transferred to a warmed 5 ml tube containing 3 ml of 2YT top agar (0.5% (w/v) agar) containing 80 µl of X-gal (2% (w/v) in DMF) and 25 µl of IPTG (25% (w/v) in H_2O). The mixture was poured instantly onto 2YT plates and allowed to set on a level surface at RT for 30 min. Plates were incubated inverted at 37°C for 24 hr. White recombinant plaques were harvested by removing agar plugs using a Pasteur pipette and stored in 200 µl of phage buffer at 4°C.

2.27.11. Preparation of M13 template

Aliquots of stored plaques were spotted onto lawns of *E. coli* TG1 using 2YT top (0.7% (w/v) agar) and bottom (1.5% (w/v) agar) agar and incubated overnight at 37°C. Using a toothpick, the zone of lysis was touched to remove a sample of cells/phage. The toothpick was then dropped into a phage tube (10 ml test tube) containing 1.5 ml of a diluted TG1 culture. This was prepared by diluting an overnight TG1 culture 1:100 in 2YT. The infected culture was allowed to grow for 6 hr with vigorous aeration (300 rpm) at 37°C using a G24 New Brunswick Scientific shaking incubator (clone grower). Cultures were then transferred to 1.5 µl microfuge tubes. After centrifugation for 5 to 10 min at high speed in a microfuge, an aliquot (1 ml) of supernatant was removed and added to an microfuge tube containing 0.25 ml of PEG/NaCl (20% (w/v) PEG/2.5 M NaCl). After mixing, the precipitation of phage was allowed to proceed for 30 min at RT. The phage was pelleted by centrifugation at 12 K for 20 min using a microfuge and PEG removed by aspiration using a yellow tip and Gilson pipette. The tube was re-spun by 'pulsing' (spinning, high speed, microfuge, 10 sec) and the remainder of the PEG removed by aspiration. The phage pellet was allowed to resuspend in 100 µl of TE for 15 min before the addition of

100 µl neutral phenol (buffered to pH 8.0, stored under TE buffer). After vigorous mixing for 1 min by vortexing, the tube was centrifuged for 5 min before removing the upper aqueous phase, taking great care so as not to remove any phenol. Template DNA was precipitated by adding sodium acetate 2.5 M (12 µl) and ethanol (300 µl) followed by storage overnight at -20°C. The tube was then centrifuged for 20 min at high speed in a microfuge and the supernatant removed by aspiration. The pellet was washed once with 100% ethanol (100 µl) which was again removed by aspiration, taking care not to dislodge the pellet. The DNA template was then resuspended in 30 µl of TE and stored at -20°C.

2.27.12. Sequencing reactions using 'Sequenase'

Sequence reactions were carried out as described by the manufacturers (United States Biochemical Corporation) except using 3 µl of template instead of 7 µl and making the annealing mix up to 10 µl using H_2O. Multiple sequencing reactions were carried out simultaneously (up to 12 at a time) using a multi-dispensing electronic Gilson pipette and micro-titre dishes.

2.27.12.1. Annealing template and primer

The following reagents were mixed in a 500 µl microfuge tube:

Universal primer	1 µl
Reaction buffer (5 x)	2 µl
DNA (template)	3-5 µl

The volume was made up to 10 µl with H_2O and the tube incubated at 60°C for 10 min using a water bath. The water bath was then switched off and allowed to cool slowly until reaching RT. The microfuge tube was then stored at 4°C until needed, and the annealed template was used within 2 hr.

2.27.12.2. Labelling reaction

Before initiating the labelling reaction, the termination mixes were dispensed into a microtitre dish (section 2.27.12.3.). The following reagents were added to the 10 µl of annealed template/primer mixture:

Template/Primer	10.0 µl
DTT (0.1 M)	1.0 µl
Labelling mix (undiluted)	2.0 µl
[^{35}S]dATP	0.5 µl

The tube was then pulse centrifuged (spun for 10 sec at high speed using a microfuge) to ensure that all the reaction mixes were at the bottom of the tube. Immediately prior to initiating the reaction, the 'Sequenase' enzyme was diluted 1:8 in ice cold 'Sequenase dilution buffer'. The diluted 'Sequenase' (2 µl) was then placed onto the side of the microfuge tube containing the rest of the reaction mix. The labelling reaction was initiated by pulse centrifuging the tube for 5 sec using a microfuge and allowed to proceed for 3-5 min at RT (23°C).

2.27.12.3. Termination reactions

After completion of the 'labelling reaction', the sample (14 µl) was dispensed (3.5 µl) onto the sides of four wells of a microtitre dish already

containing each of the four termination mixes (A, C, G, T). The microtitre dish was then centrifuged until reaching 2000 rpm using a MSE Mistral 2000 to initiate the termination reactions, and then incubated at 37°C for 10 min. After this time, 4 µl of 'stop solution' was added to the side of each well. All reactions were stopped simultaneously by centrifuging the microtitre dish until a speed of 2000 rpm was reached. Samples were then either frozen at -20°C or used immediately. Prior to loading the samples onto the gel, they were heated at 80°C for 15-25 min or until they evaporated to a volume of 2-4 µl.

2.27.12.4. Multiple sequencing reactions

Using this method it was possible to sequence up to 12 templates at a time. It was essential that diluted 'Sequenase' was dispensed with speed onto the sides of the microfuge tubes during the labelling reaction using an electronic multi-dispensing Gilson pipette. After pulse centrifuging the tubes to initiate this reaction, the samples were immediately dispensed to the sides of the wells containing the termination mixes, again using the electronic multi-dispensing Gilson pipette. The termination reactions were initiated (i.e. the microtitre dish was centrifuged) immediately after the transfer of the labelling reactions to the microtitre dish was complete (< 4min). The quality of the sequence was often dependent upon the time taken to initiate the termination reaction, and deteriorated if the labelling step exceeded 5 min.

Solutions for sequencing with 'Sequenase'

Solutions were either made from lab stocks or purchased as a kit with the 'Sequenase' enzyme.

'Sequenase' reaction buffer (5 x)

200 mM Tris.HCl pH 7.5

100 mM $MgCl_2$

250 mM NaCl

Universal (-40) primer

5'-GTTTTCCCAGTCACGAC-3' 0.5 pM/µl

Labelling mix (dGTP)

7.5 µM dGTP, 7.5 µM dCTP, 7.5 µM dTTP

This solution was used neat and not diluted as recommended by the Manufacturers.

Labelling mix (dITP)

15 µM dITP, 7.5 µM dCTP, 7.5 µM dTTP

This solution was used neat and not diluted as recommended by the Manufacturers.

ddG Termination mix (for dGTP)

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP, 50 mM NaCl

ddA Termination mix (for dGTP)

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddATP, 50 mM NaCl

ddT Termination mix (for dGTP)

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddTTP, 50 mM NaCl

ddC Termination mix (for dGTP)

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddCTP, 50 mM NaCl

Sequence extension mix (for dGTP)

180 μM each of dGTP, dATP, dCTP, dTTP

ddG Termination mix (for dITP)

160 μM dITP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 1.6 μM ddGTP, 50 mM NaCl

ddA Termination mix (for dITP)

80 μM dITP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddATP, 50 mM NaCl

ddT Termination mix (for dITP)

80 μM dITP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddTTP, 50 mM NaCl

ddC Termination mix (for dITP)

80 μM dITP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddCTP, 50 mM NaCl

Sequence extending mix (for dITP)

360 μM dITP, 180 μM each of dATP, dCTP, dTTP

Enzyme dilution buffer

10 mM Tris.HCl pH 7.5

5 mM DTT

0.5 mg/ml BSA

Stop solution

95% (v/v) Formamide

20 mM EDTA

0.05% (w/v) Bromophenol blue

0.05% (w/v) Xylene Cyanol FF

2.27.13. Plasmid sequencing

Plasmid sequencing was performed according to S. van der Vies (pers. comm.) using denatured midi-prep plasmid DNA (section 2.18.3.) as a template. This method has been described by Murphy and Kavanagh (1988).

Sepharose columns were prepared using 1.5 ml microfuge tubes and 0.7 ml microfuge tubes, both with their lids removed. The base of a 0.7 ml microfuge tube was pierced using the tip of a 23 gauge (blue) syringe needle which was then removed. A 1.5 ml microfuge tube was also pierced using the same needle but this time the needle was pushed completely in before removal. The 0.7 ml microfuge tube was placed inside the 1.5 ml microfuge so it was resting on the collar. The 1.5 ml microfuge tube was then positioned in the same way in a 5 ml plastic centrifuge tube. One drop of glass beads (Ballotini no. 1 acid washed) was then placed in the bottom of

the 0.7 ml microfuge tube which was then filled with a 66% 'Sepharose' CL-6B' slurry in $TE_{0.1}$ buffer (Tris.HCl 10 mM, EDTA 0.1 mM pH 8.0). This slurry had been autoclaved and stored at 4°C. The plastic centrifuge tube (containing the two microfuge tubes) was then centrifuged in a swing-out rotor (Radwell H-103N centrifuge) at 3 K for 3 min in order to pack the spin-column. The 1.5 ml microfuge tube was then replaced with an unpierced one. At this stage plasmid DNA (20 µl) was denatured by the addition of 1 µl of 5 M NaOH and incubation at RT for exactly 5 min. The denatured DNA was immediately transferred to the top of the 'Sepharose' spin-column and centrifuged at 3 K for 3 min. The DNA (16 µl recovered) was collected from the bottom of the 1.5 ml microfuge tube and stored at 4°C until required. This template was sequenced using the protocol described in section 2.27.12. except using 7 µl of template in the annealing step.

2.27.14. Sequencing gels

Gradient gels (1-5%) were poured using 50 x 20 cm plates which had been thoroughly cleaned using detergent, EtOH and finally acetone. The back plate was siliconised using 'repelcote' (from BDH), allowed to dry and then cleaned again using EtOH. Sequencing gel plates were assembled according to the manufacturers instructions. Gradient gels were poured as follows. Two beakers, one containing 7 ml of 5 x acrylamide solution, the other containing 50 ml of 1 x acrylamide solution were obtained. Polymerisation was initiated by the addition of APS (25% (w/v)) and TEMED as described below.

7 ml 5 x acrylamide	50 ml 1 x acrylamide
14 µl APS	100 µl APS
14 µl TEMED	100 µl TEMED

A 50 ml syringe was used to withdraw 40 ml of the 1 x acrylamide and then was then placed aside. A 10 ml pipette was used to withdraw 5 ml of the remaining 1 x acrylamide mix. With the 1 x acrylamide mix still in the 10 ml pipette, 5 ml of the 5 x acrylamide mix was taken up. An air bubble was introduced into the pipette to mix the two solutions. The contents of the 10 ml pipette were then poured (slowly but with constant flow) between the two glass plates of the assembled gel receptacle. The plates were held at an angle of 30° to the horizontal. The flow of the gel mix was directed down one side by tilting the receptacle accordingly. After releasing the contents of the 10 ml pipette, the remaining 1 x acrylamide mix was poured (without breaking the flow of solution) using the syringe. When the receptacle was full, the well-forming comb (wiped with APS) was positioned near the top of the gel. The plates were then clamped tightly together using clips. Once set (typically 30 min), the well-forming comb was removed and the wells rinsed thoroughly with running buffer (1 x TBE). Gels were placed into gel tanks and upper and lower buffer reservoirs were filled with running buffer (1 x TBE). Samples were loaded using a 3 µl Laser pipette (from Drummond Laboratory Systems) and run at 2000 V until the leading dye front had reached the bottom of the gel (approx. 3½ hr). Gels were then fixed in 10% (v/v) acetic acid and dried under vacuum at 80°C for 20-30 min using a gel drier (Biorad). Gels were then subjected to autoradiography for 24-48 hr at RT.

Gel solutions

10 x TBE buffer (1 l)

Tris	109 g
boric acid	55 g
EDTA	9.3 g

40% Acrylamide stock solution (100 ml)

acrylamide	38 g
bis-acrylamide	2 g

This solution was de-ionised by stirring over 2 g amberlite resin (BDH) for 30 min, filtered and stored at 4°C in the dark.

1 x TBE acrylamide urea gel mix (per l)

urea	430 g
10 x TBE	100 ml
40% acrylamide/bis	150 ml

5 x TBE acrylamide urea gel mix (per l)

urea	430 g
10 x TBE	500 ml
40% acrylamide/bis	150 ml
BPB	30 mg

All acrylamide solutions were stored at 4°C in the dark.

2.28. Identification of proteins

2.28.1. In vitro coupled transcription/translation 'Zubay' system

The bacterial cell-free coupled transcription/translation system allows the expression of genes contained on a plasmid (Zubay, 1973). The gene of interest must, however, have the necessary start signals for transcription and translation. Proteins produced by this system can be labelled using radioactive ^{35}SMet and visualised after SDS PAGE by autoradiography.

The 'Zubay' in vitro coupled transcription/translation system was obtained as a kit from Amersham International. The procedure was carried out exactly as described by the manufacturers instructions. Samples were then subjected to SDS PAGE and visualised by autoradiography. Radioactive 'Rainbow' molecular weight markers, also obtained from Amersham International, were run on the gel in order to determine the size of any proteins obtained.

2.28.2. T7 RNA polymerase directed expression system

This method allows the controlled expression of genes under the control of T7 promoters. The protocol has been described by Tabor and Richardson (1985).

2.28.2.1. Construction of recombinant plasmids and strains

The plasmids used for cloning were pT7-5 or pT7-6 (see Appendix I). A multiple cloning site (MCS) was situated directly in front of the T7 RNA polymerase promoter. DNA fragments were cloned into the MCS using standard molecular biology techniques (See sections 2.18. to 2.26.).

After constructing the appropriate plasmids and checking them in

E. coli DHI, they were then introduced into E. coli K38 (containing pGP1-2 which carries the gene encoding T7 RNA polymerase) by transformation. Transformants were selected by their ability to grow on NA plates containing Kn (to maintain pGP1-2) and Ap (to select pT7-5 or pT7-6 recombinant derivatives). It was important to select these E. coli transformants at 30°C to ensure that the temperature control over the expression of T7 RNA polymerase was not lost.

2.28.2.2. Expression of cloned genes

A fresh E. coli K38 colony containing the two plasmids (pGP1-2 and pT7-5/6 derivative) was grown overnight at 30°C in LB containing Ap and Kn. This culture was then diluted 1:40 in the same medium (5 ml) and transferred to a 50 ml conical flask. The culture was incubated with shaking (200 rpm) at 30°C until reaching A_{590} = 0.4. At this stage 1 ml of culture was removed and centrifuged (high speed, microfuge for 2 min) and the pellet was washed once with M9 medium (5 ml). The pellet was then resuspended in 5ml of M9 medium which had been supplemented with 18 amino acids (lacking cysteine and methionine) and transferred to a 50 ml flask. This culture was then incubated with shaking (200 rpm) at 30°C for 30 min. The culture was induced by transferring the flask to a shaking water bath (200 rpm) set at 42°C for 20 min. Rifampicin was added to a final concentration of 200 µg/ml (from a stock of 20 mg/ml dissolved in methanol and stored at -20°C). The cells were incubated for a further 10 min at 42°C before returning to the 30°C shaking water bath. After 20 min incubation at 30°C, an aliquot (0.5 ml) of cells was removed and transferred to a 2 ml microfuge tube containing 10 uCi of ^{35}S-methionine. The culture was pulsed with this radio-labelled amino acid by incubating the tubes in a New

Brunswick Scientific G24 shaking incubator for 5 min at 30°C and 200 rpm. The cells were then pelleted by centrifugation for 1 min at high speed in a microfuge. The cell pellet was then resuspended in 150 µl of 1 x cracking buffer (60 mM Tris.HCl pH 6.8, 1% 2-mercaptoethanol, 10% (v/v) glycerol and 0.01% (w/v) bromophenol blue) and denatured by incubating in a boiling water bath for 5 min. Samples were frozen at -20°C for storage or loaded (40 µl) onto SDS polyacrylamide gels. Gels were electrophoresed overnight and then treated with 'Amplify' (Amersham International) according to the manufacturers instructions. Gels were then dried under a vacuum at 60°C for 2 hr and then subjected to autoradiography at RT overnight.

2.29. SDS polyacrylamide gel electrophoresis (SDS PAGE)

Proteins were analysed on 10% slab gels as described by Laemmli (1970). Gels were poured using a 'Biorad' 50 ml vertical gel apparatus. The apparatus was assembled as described by the manufacturers. The 10% gel mix was poured slowly between the glass plates using a 25 ml pipette until approximately 5 cm from the top of the plates. Butanol (H_2O saturated) was then poured onto the top of the gel mix in order to ensure that the top of the gel was level when polymerisation was complete (1-2 hr). The butanol was then washed off by rinsing with H_2O before adding the stacking gel mix. The stacking gel mix was added until it filled the gap between the top of the running gel and the top of the glass plates. The well-forming comb was instantly inserted into the stacking gel leaving a space of 3 cm between the top of the running gel and the bottom of the wells. When set (20 min), the comb was carefully removed and the wells washed out with running buffer using a Pasteur pipette and teat. The gel was then transferred to the gel

tank containing running buffer and the samples loaded through the upper buffer chamber.

10% lower gel mix (50 ml)	Stacking gel (10 ml)
8.3 ml hi-bisacrylamide	3.0 ml stacking gel acrylamide
34.9 ml H_2O	2.2 ml H_2O
6.25 ml lower gel buffer	2.4 ml stacking gel buffer
0.5 ml 10% SDS	0.1 ml SDS
[10 µl TEMED]	[5 µl TEMED]
[100 µl APS]	[100 µl APS]

Items in [brackets] were added after de-gassing the solution (20 min under vacuum).

Stacking gel buffer

Tris (0.5 M) was made by dissolving 5.98 g in 80 ml of H_2O and then adjusted to pH 6.8 before being made up to 100 ml with H_2O.

Lower gel buffer

Tris (0.5 M) was made by dissolving 36.33 g in 80 ml of H_2O and then adjusted to pH 8.8 before being made up to 100 ml with H_2O.

Hi-bis acrylamide (low % gel) mix

A 60% stock was made by dissolving 60 g of acrylamide and 1.6 g of bis-acrylamide in 25 ml of H_2O. The volume was later made up to 100 ml with H_2O. This stock was stored at 4°C.

Stacking gel acrylamide

10 g of acrylamide and 0.5 g of bis-acrylamide were dissolved in a final volume of H_2O of 100 ml. This stock was stored at 4°C.

Running buffer

A 5 x concentrated stock was made by dissolving 30.2 g of Tris and 144 g of glycine in H_2O in a final volume of 1 l. Running buffer was prepared by mixing 200 ml of this 5 x stock with 10 ml of 10% SDS and 780 ml of H_2O.

Stacking gel acrylamide

10 g of acrylamide and 0.5 g of bis-acrylamide were dissolved in a final volume of H_2O of 100 ml. This stock was stored at 4°C.

Running buffer

A 5 x concentrated stock was made by dissolving 30.2 g of Tris and 144 g of glycine in H_2O in a final volume of 1 l. Running buffer was prepared by mixing 200 ml of this 5 x stock with 10 ml of 10% SDS and 780 ml of H_2O.

CHAPTER 3

THE SEARCH FOR *Ecc* MUTANTS DEFECTIVE IN EXTRACELLULAR ENZYME PRODUCTION

3.1. Introduction

The generation of *Erwinia* spp. mutants defective in extracellular enzyme production has been described previously (section 1.9.8.). Various techniques have been used to generate such mutants including transposon and chemical mutagenesis. Mutants of *Ecc* defective in extracellular enzyme production have been generated using various transposons (Tn*5*, Tn*phoA* and Tn*10*) in this laboratory (unpublished). However, in this study the chemical mutagen EMS was used. A brief discussion on the merits of transposon and chemical mutagenesis will be given. This will be followed by a description of the approach taken in this study and a discussion on the classes of *Ecc* mutants isolated.

3.2. Mutagenesis - Transposons versus chemical mutagens

Transposons can be used to generate single and tagged mutations. This is a useful feature for cloning mutated genes. However, a major disadvantage is that transposons sometimes integrate into 'preferred sites' or 'hot-spots' in the bacterial chromosome. This may lead to the inability to generate mutations in some genes. Mutations generated by transposons are caused by an insertion event into a gene resulting in a gross disruption of that gene. It was not known at the onset of this study if transposon insertions into genes encoding extracellular enzyme secretion proteins would be lethal. The lethality of mutations in the genes involved in the export apparatus of *E. coli* (*sec* genes) has been discussed (section 1.4.2.) and there was no reason to suggest that the secretion apparatus might not also be part of a general, essential process.

Chemical mutagens can be used to generate subtle and random mutations by causing single base changes. Point mutations sometimes result in

conditional mutants which display a phenotype dependent on external factors such as temperature. This technique has been used to isolate mutations in essential processes including cell division and protein export in E. coli. Mutants conditionally defective in extracellular enzyme production were searched for in this study.

There are some disadvantages when using chemical mutagens. One problem is the generation of multiple mutations. This can be reduced by optimising the conditions of mutagenesis to produce a low but detectable number of mutants. Mutations produced by chemicals (or UV light) cannot be easily cloned, unlike transposon insertion mutants. Furthermore, chemical mutagens are often highly toxic and carcinogenic.

3.3. The use of ethyl methyl sulphonate (EMS) to generate mutants of Ecc

In this study the chemical mutagen ethyl methyl sulphonate (EMS) was used. The chemical formula of the EMS molecule is $CH_3SO_3CH_2CH_3$. This chemical is a powerful alkylating agent and has been used previously to generate mutations in Ecc SCRI193 (Forbes and Perombelon, 1985). EMS alkylates the purine bases, adenine and guanine, which are subsequently lost from the DNA (phosphodiester linked deoxyribose) backbone. During the next round of DNA replication, DNA synthesised from this damaged template has random nucleotides inserted in the positions where gaps in the template exist (Stanier et al., 1980).

3.4. Results

3.4.1. Construction of a killing curve

Ecc HC131 was grown and treated with EMS as described in section 2.6.

Samples of bacteria were taken at 30 min intervals after the addition of EMS. The viable count was determined at each time point and the percentage survival as compared with the viable count before the addition of EMS (t=0 min) was determined.

On the first run of this experiment, an exposure to EMS for 180 min resulted in a survival of 0.4%. The killing curve was then repeated taking frequent culture samples at 20 min intervals from between 100 and 220 min after the addition of EMS. The results are illustrated in Figure 3.1.

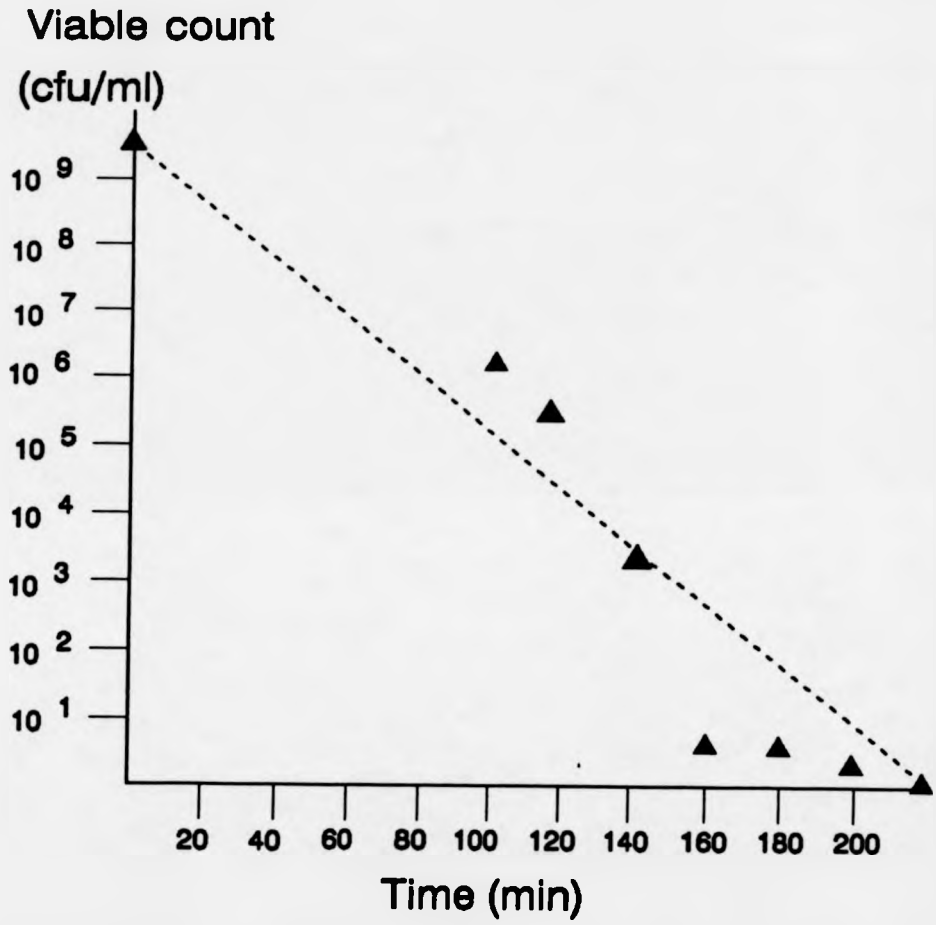
A survival of between 1 and 4% was sought which was expected to generate auxotrophs in the surviving culture at a frequency of 1 to 4% (Forbes and Perombelon,1985).

3.5. Survival of Ecc after EMS treatment

A culture of Ecc was treated as described in section 2.6. under conditions expected cause adequate mutagenesis, as determined from the killing curve (Figure 3.1.). The survival of a bacterial culture of Ecc after being treated with EMS for 90 min was found to be 3.5%. The survival is an expression of the viability of the culture after EMS treatment (t=90 min) compared to the viability before (t=0 min).

This frequency of survival was within the limits used by Forbes and Perombelon (1985). The mutagenised bacterial suspension was stored at 4°C until required. The mutagenised cell suspension was serially diluted and plated

Figure 3.1. Survival of *Ecc* treated with EMS



Key

▲ Sample point

cfu Colony forming units

out to give approximately 500 to 1000 colonies per plate.

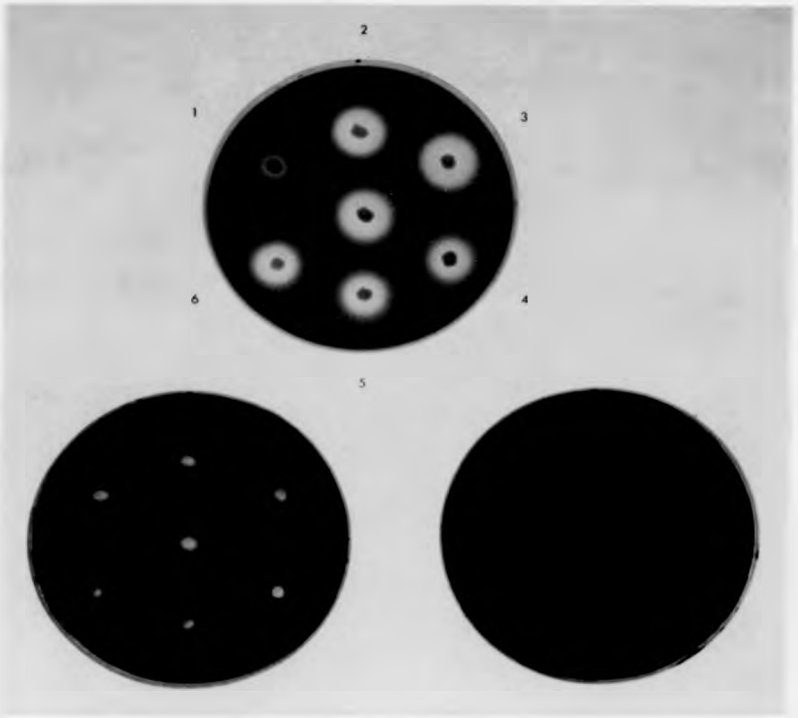
In order to determine the proportion of auxotrophs in the mutagenised population, 100 colonies were screened on NA and MMA plates. From screening this small number of colonies, two auxotrophs were found. Upon screening a further 504 colonies on NA, MMA and also extracellular enzyme detection plates, a further 14 auxotrophs were found, a frequency of 2.7%.

The mutagenised stock was screened as described in section 2.9. The mutagenised stock was stored as colonies on NA Ap plates and as a bacterial suspension, in both cases at 4°C. After screening 7000 mutagenised colonies on extracellular enzyme plates, MMA and NA plates, a variety of mutant phenotypes were discovered. Some of these mutant classes will be described in the following section.

3.6. Classes of Ecc mutants generated by treatment with EMS

Figure 3.2. is a photograph showing some of the Ecc mutants defective in the ability to produce extracellular enzymes. In the centre of each developed plate is the parent strain Ecc HC131. The halos surrounding the bacterial colonies indicate the production of extracellular enzymes. Surrounding the wild-type strain are six mutants identified by their altered halo sizes (altered levels of enzyme production). Colony 1 is reduced for Prt production and the levels of production for both Pel and Cel also appear to be reduced. Colonies 2 and 6 produce Prt but are reduced in the production of both Pel and Cel. Colony 3 appears to hyper-produce all three classes of extracellular enzyme whereas colony 5 hyper-produces Cel but is unaffected for Prt and Pel production. Colony 4 was isolated on the basis of having an altered halo on the Pel plate. The inner halo was lacking but the outer halo was still present. The different phenotypic classes of mutants isolated and

Figure 3.2. EMS mutants of HC131 altered in extracellular enzyme production



Key

Grey/green plate	Protease (Prt) detection
Dark blue plate	Cellulase (Cel) detection
Turquoise plate	Pectinase (Pel) detection

Colony position	Strain	Phenotype
Centre	HC131 (parent)	Pel+ Cel+ Prt+
1	RJP113	Pel+ Cel- Prt- (Sex-)
2	RJP253	Pel- Cel- Prt+ (Out-)
3	RJP133	Pel++ Cel++ Prt++ (hyper)
4	RJP241	Pel? Cel+Prt+
5	RJP209	Pel+ Cel++ Prt+ (Cel hyper)
6	RJP233	Pel- Cel- Prt+ (Out-)

Table 3.1. Classes of extracellular enzyme mutants generated using EMS

Mutant	Number	Frequency (%)
Auxotrophic	136	2.4
Pel+, Cel+, Prt-	43(35)	(0.6)
Pel+, Cel-, Prt-	38	0.7
Pel-, Cel-, Prt+	14	0.2
Prt^{ts}	7	0.1
Pel-, Cel-, Prt-	6	0.1
Pel+, Cel-, Prt+	3	0.05
Pel?, Cel+, Prt+	3	0.05
Pel++, Cel++, Prt++ (hyper all 3 enzymes)	1	0.02
Pel+, Cel++, Prt+ (Cel hyper)	1	0.02

Legend

A total of 5710 mutagenised colonies were screened at two temperatures (25°C and 33°C). Prt^{ts} mutants produced Prt at 25°C but were negative for this enzyme (or greatly reduced) at 33°C. All putative mutant colonies were re-screened in order to confirm enzyme phenotypes. Some Prt- mutants (8) were obtained by plating the mutagenised bacterial stock directly onto 'skimmed-milk' (SMNA) non-destructive protease assay plates. These mutants were deducted from this class in order to calculate the true frequency of this particular mutation as indicated in parentheses.

Key

+ = positive for enzyme
- = negative for enzyme
? = altered for enzyme
++ = hyper-producer of enzyme

the frequencies at which each class arose are shown in Table 3.1. A large number of mutants demonstrating altered extracellular enzyme phenotypes were isolated. The requirements of the auxotrophic mutants were not determined. A catalogue of all the mutants isolated is given in Table 3.2.

3.7. Discussion

The range of different classes of mutants isolated from this mutagenesis procedure indicated that the mutagenesis had been random. It is interesting that the frequency of each class of mutant is different. Some mutations such as [Prt-, Pel+, Cel+] and [Cel-, Prt-, Pel+] arose at relatively high frequencies whereas others, such as [Cel-, Pel+, Prt+], were rarely isolated. The absence of other classes of mutations which might have been expected is also of interest. Most notable is the absence of any mutants singularly affected in Pel production.

As was demonstrated in Figure 3.2. and Table 3.1., a variety of classes of mutants have been generated with altered extracellular enzyme phenotypes. Mutant colonies exhibiting an altered extracellular enzyme phenotype could be the result of different events at the molecular level. Mutations in the gene encoding the actual extracellular enzyme may inactivate this enzyme, leading to the loss of this particular activity. Also, mutations in genes involved in the regulation or secretion of the extracellular enzyme might be expected to result in a similar phenotype. It is therefore necessary to distinguish between these types of mutations.

The classes of mutant fall into three main categories: the hyper-producers, those defective in the ability to produce one enzyme whilst still producing wild-type levels of other extracellular enzymes, and lastly, mutants pleiotropically defective in the ability to produce two or more classes of

Legend for Table 3.2.

Key

+	produces extracellular enzyme
-	does not produce extracellular enzyme
++	hyperproducer of extracellular enzyme
?	altered halo on extracellular enzyme detection plate

All Ecc phenotypes were determined using extracellular enzyme plates. The Out- mutants [Pel-, Cel-, Prt+] are also shown in Table 2.1. The mutant classes are discussed in the text.

Table 3.2. Extracellular enzyme mutants of Ecc HC131 generated using EMS

Strain	Total	Phenotype	
RJP122		Out-	[Pel-,Cel-,Prt+]
RJP159		Out-	
RJP190		Out-	
RJP200		Out-	
RJP208		Out-	
RJP211		Out-	
RJP220		Out-	
RJP221		Out-	
RJP233		Out-	
RJP249		Out-	
RJP250		Out-	
RJP251		Out-	
RJP253		Out-	
RJP254		Out-	
	(14)		
RJP109		Sex-	[Cel-, Prt-, Pel+]
RJP110		Sex-	
RJP111		Sex-	
RJP112		Sex-	
RJP113		Sex-	
RJP114		Sex-	
RJP116		Sex-	
RJP117		Sex-	
RJP118		Sex-	
RJP119		Sex-	
RJP120		Sex-	
RJP122		Sex-	
RJP123		Sex-	
RJP128		Sex-	
RJP160		Sex-	
RJP161		Sex-	
RJP181		Sex-	
RJP188		Sex-	
RJP189		Sex-	
RJP191		Sex-	
RJP192		Sex-	
RJP196		Sex-	
RJP203		Sex-	
RJP204		Sex-	
RJP205		Sex-	
RJP206		Sex-	
RJP212		Sex-	
RJP222		Sex-	
RJP223		Sex-	
RJP226		Sex-	

Table 3.2. (cont.)

Strain	Total	Characteristics
RJP236		Sex-
RJP239		Sex-
RJP240		Sex-
RJP243		Sex-
RJP246		Sex-
RJP247		Sex-
RJP256		Sex-
RJP260		Sex-
	(38)	
RJP115		[Pel-, Cel-, Prt-]
RJP124		[Pel-, Cel-, Prt-]
RJP187		[Pel-, Cel-, Prt-]
RJP193		[Pel-, Cel-, Prt-]
RJP248		[Pel-, Cel-, Prt-]
RJP259		[Pel-, Cel-, Prt-]
	(6)	
RJP103		[Prt-, Pel+, Cel+]
RJP106		[Prt-, Pel+, Cel+]
RJP107		[Prt-, Pel+, Cel+]
RJP126		[Prt-, Pel+, Cel+]
RJP127		[Prt-, Pel+, Cel+]
RJP134		[Prt-, Pel+, Cel+]
RJP135		[Prt-, Pel+, Cel+]
RJP137		[Prt-, Pel+, Cel+]
RJP139		[Prt-, Pel+, Cel+]
RJP140		[Prt-, Pel+, Cel+]
RJP142		[Prt-, Pel+, Cel+]
RJP143		[Prt-, Pel+, Cel+]
RJP145		[Prt-, Pel+, Cel+]
RJP147		[Prt-, Pel+, Cel+]
RJP148		[Prt-, Pel+, Cel+]
RJP149		[Prt-, Pel+, Cel+]
RJP163		[Prt-, Pel+, Cel+]
RJP164		[Prt-, Pel+, Cel+]
RJP165		[Prt-, Pel+, Cel+]
RJP166		[Prt-, Pel+, Cel+]
RJP186		[Prt-, Pel+, Cel+]
RJP197		[Prt-, Pel+, Cel+]
RJP198		[Prt-, Pel+, Cel+]
RJP201		[Prt-, Pel+, Cel+]
RJP202		[Prt-, Pel+, Cel+]
RJP207		[Prt-, Pel+, Cel+]
RJP210		[Prt-, Pel+, Cel+]
RJP215		[Prt-, Pel+, Cel+]

Table 3.2.(cont.)

Strain	Total	Characteristics
RJP216		[Prt-, Pel+, Cel;+]
RJP217		[Prt-, Pel+, Cel+]
RJP219		[Prt-, Pel+, Cel+]
RJP225		[Prt-, Pel+, Cel+]
RJP227		[Prt-, Pel+, Cel+]
RJP228		[Prt-, Pel+, Cel+]
RJP229		[Prt-, Pel+, Cel+]
RJP234		[Prt-, Pel+, Cel+]
RJP235		[Prt-, Pel+, Cel+]
RJP237		[Prt-, Pel+, Cel+]
RJP238		[Prt-, Pel+, Cel+]
RJP245		[Prt-, Pel+, Cel+]
RJP252		[Prt-, Pel+, Cel+]
RJP258		[Prt-, Pel+, Cel+]
RJP262		[Prt-, Pel+, Cel+]
	(43)	
RJP131		Prt^{ts} [Prt+, 25C, Prt- 33C]
RJP132		Prt^{ts}
RJP136		Prt^{ts}
RJP141		Prt^{ts}
RJP146		Prt^{ts}
RJP163		Prt^{ts}
RJP218		Prt^{ts}
	(7)	
RJP232		[Cel-, Pel+, Prt+]
RJP213		[Cel-, Pel+, Prt+]
RJP214		[Cel-, Pel+, Prt+]
	(3)	
RJP133		[Prt++, Cel++, Pel++] [hyperproducer]
	(1)	
RJP209		[Prt+, Cel++, Pel+] [Cel hyperproducer]
	(1)	
RJP224		Cel+, Prt+, Pel fuzzy (Pel?)
RJP241		Cel+, Prt+, Pel fuzzy (Pel?)
RJP261		Cel+, Prt+, Pel fuzzy (Pel?)
	(3)	

extracellular enzyme.

Only two hyper-producing mutants were generated as determined by the increased size of halo on the enzyme detection plate. One mutant [Cel++, Prt+, Pel+] had enhanced Cel activity. This could be due to higher activity of the Cel enzyme i.e. an enhanced catalytic activity mutant. Alternatively, there could be higher amounts of Cel production or enhanced secretion of this enzyme. The other hyper-secretion mutant [Pel++, Cel++, Prt++] produced larger halos on all extracellular enzyme detection plates. It is unlikely that this phenotype arose by three separate mutation events leading to increased catalytic activity of all three enzymes. A much more plausible explanation is that a common regulation or secretion mechanism had been mutated.

The second class of mutant, those which displayed an alteration in one extracellular enzyme, could have arisen as a result of mutations in the enzyme itself. An alternative explanation is that these mutants are defective in the regulation or secretion of that particular enzyme. Mutants of this class include those with [Cel-, Prt+, Pel+] and [Cel+, Prt- Pel+] phenotypes. Mutant strains with a [Pel-, Cel+, Prt+] phenotype were never detected. However, it is known that there are several isozymic forms of this enzyme (section 1.9.5.5.1.). This would make it impossible to detect a mutation in any individual pel gene. The Cel- mutants have recently been complemented with a plasmid containing the cellulase structural gene (celV) from Ecc SCRI193 (V. Cooper, pers. comm.). This result strongly suggests that the [Cel-, Prt+, Pel+] mutants had cel structural gene mutations.

The third class of mutants are those pleiotropically defective in the ability to produce at least two types of enzyme. The two subclasses which were discovered displayed [Pel-, Cel- Prt+] or [Cel+, Pel-, Prt-] phenotypes. Mutants of this class must be defective in either regulation or secretion of

the two extracellular enzymes involved. The [Cel+, Pel-, Cel-] mutants were not studied further in this study. However, other workers have demonstrated that these mutants have lowered levels of synthesis of Cel, Pel and Prt (S. Stevens, pers. comm.). This phenotype is now known as Sex- (Synthesis of extracellular enzymes).

The [Pel-, Cel-, Prt+] mutants (designated Out-) have been introduced in section 1.9.8.). The following chapters in this thesis will describe the characterisation of this class of mutant.

CHAPTER 4

CHARACTERISATION OF MUTANTS WITH THE [Pel-, Cel-, Prt+] PHENOTYPE: THE SEARCH FOR SECRETION (Out-) MUTANTS

4.1. Introduction

A wide variety of mutants defective in extracellular enzyme production were generated using the chemical mutagen EMS (section 2.6.). However, for the purposes of this study, only mutants truly defective in the ability to secrete extracellular enzymes were considered to be of importance. Other workers (Andro *et al.*, 1984; Ji *et al.*, 1987, 1989; Thurn and Chatterjee, 1985; Murata *et al.*, 1990) have generated *Ecc* and *Ech* mutants defective in extracellular enzyme production and have demonstrated that those with the [Pel-, Cel-, Prt+] phenotype accumulated Pel and Cel intracellularly. Mutants with the [Pel-, Cel-, Prt+] phenotype will be called Out- mutants which is consistent with their description in the scientific literature.

This chapter will concentrate on characterising the Out- mutants generated from this work and determining whether or not they are truly defective in secretion or are altered in regulation. Mutants defective in extracellular enzyme secretion, but unaltered in extracellular enzyme production, would be expected to accumulate the normally extracellular enzyme(s) intracellularly. This assumes that there are no feedback control mechanisms which may 'switch off' enzyme synthesis in response to such an accumulation. By growing a culture of each mutant in enzyme induction medium and then lysing the cells, it is possible to determine whether or not the extracellular enzymes are still being synthesised. This is achieved by assaying the levels of extracellular enzymes in the culture supernatant and in the cell sonicate.

4.2. Do the extracellular enzymes Pel and Cel accumulate intracellularly in [Pel-, Cel-, Prt+] Out- mutants ?

Cultures of Out- mutants and the wild-type strain (HC131) were grown

in extracellular enzyme inducing medium (PM) and supernatant and sonicate fractions were prepared as described in section 2.16. All cultures were harvested at similar cell densities (A_{600}=2.5-3.0) after overnight growth (14-16 hr). The cell fractions were then subjected to enzyme assays to determine levels of Pel (section 2.10.3.), Bla (section 2.10.2.) and Cel (section 2.10.6.1.) activity. The levels of protease activity were not determined for this preliminary investigation but appeared to be similar to those of the wild-type parent strain as determined by the qualitative plate assays.

4.2.1. Results

The results for this experiment are summarised in Table 4.1. The parent strain Ecc HC131 was used as the wild-type control and the mutants are listed beneath. The parent strain, which was used as the standard for extracellular enzyme production levels, had a typical distribution of enzymes with Pel and Cel being predominantly extracellular (92% and 78% respectively). The 14 Out- mutants accumulated Pel (>97%) and Cel (>90%) intracellularly. Low levels of activity of these enzymes (<5%) were present in the cell supernatant. The marker enzyme Bla was correctly localised within the cellular fraction (>96%) with a small amount (<4%) leaking into the supernatant.

4.2.2. Discussion

These findings are consistent with the results from the plate assays in showing that greatly reduced levels of Pel and Cel are reaching the extracellular milieu in Out- mutants. However, this experiment also demonstrated that Pel and Cel were still produced but remained within the cell. When compared with the parent strain, all the mutants of this class had

Table 4.1. Enzyme localisations in Out- mutants

	Enzyme								
	Pel			Cel			Bla		
	Enzyme acivity (%) in fraction								
Strain	SO	SN	%wt	SO	SN	%wt	SO	SN	%wt
HC131	8	92	100	22	78	100	97	3	100
RJP233	98	2	58	100	0	9	97	3	105
RJP200	99	1	121	100	0	22	97	3	82
RJP254	97	3	275	90	10	37	96	4	93
RJP249	99	1	109	100	0	28	98	2	84
RJP190	97	3	154	100	0	22	98	2	98
RJP122	99	1	173	100	0	33	98	2	91
RJP211	99	1	58	100	0	33	97	3	87
RJP220	100	0	125	100	0	51	99	1	119
RJP251	99	1	39	100	0	78	97	3	110
RJP250	99	1	201	100	0	50	97	3	116
RJP208	99	1	145	100	0	77	97	3	85
RJP159	99	1	144	100	0	25	97	3	95
RJP221	98	2	44	100	0	9	99	1	95
RJP253	99	1	90	100	0	22	98	2	84

Key

SO = cell sonicate fraction

SN = cell supernatant fraction

%wt= total enzyme activity of mutant compared to the wild type (parent) strain

elevated levels of Pel and Cel within the cell. However, the levels of Pel and Cel synthesised did appear to vary between individual mutants. For example, mutant RJP221 appears to be reduced in the levels of synthesis for both Pel and Cel compared to the parent strain HC131. Other mutants, however, exhibited enhanced levels of activity (for example RJP254) particularly with respect to Pel activity. These differences between mutants might be attributable to enzyme assay irreproducibility, a problem which could be overcome by assaying multiple samples several times over. An alternative explanation is that there are phenotypic divisions (subgroups) of mutant classes within the Out- mutants. Other workers have also found that the levels of Pel and Cel synthesis do vary between Ecc mutants of this class (Chatterjee, et al., 1985a). The general trend, however, is clear in that all mutants of this class do appear to accumulate Pel and Cel intracellularly.

4.3. Identifying the location of extracellular enzyme accumulation in Out- mutants: cytoplasmic or periplasmic ?

In section 4.2. it was shown that both Pel and Cel accumulate within the cell in this particular class of mutant. In order to determine the exact cellular site of accumulation of extracellular enzymes, cultures of each mutant were grown in extracellular enzyme inducing medium and fractionated as described in section 2.17. These procedures were used to fractionate the cells in order to obtain the cytoplasmic fraction and the periplasmic fraction. The distribution of extracellular enzymes after cellular fractionation by cold osmotic shock (section 2.17.1.) and spheroplast formation (section 2.17.2.) will be presented in sections 4.3.1. and 4.3.2.

4.3.1. Fractionation by osmotic cold shock

Enzyme localisations determined for cultures fractionated using the osmotic cold shock procedure are shown in Table 4.2. Enzymes were assayed as in section 4.2. except that Cel was assayed as in section 2.10.6.3. Additionally, assays for polygalacturonase (Peh; section 2.10.5.) and β-galactosidase (β-gal; section 2.10.1.) were performed. The percentage of each enzyme found to reside in each fraction is given. The parent strain (HC131) had a typical distribution of marker enzymes. The cytoplasmic marker, β-galactosidase (β-gal), was predominantly cytoplasmic (81%) and β-lactamase (Bla), the periplasmic marker, was predominantly periplasmic (88%). The extracellular enzyme Pel was predominantly extracellular. The secretion mutant RJP237 was fractionated and assayed along with the parent strain HC131. The marker enzymes β-Gal and Bla were in their appropriate compartments ie. cytoplasm and periplasm respectively. The normally extracellular enzyme Pel was predominantly cytoplasmic (64.42%). However, there was a significant amount of activity in the periplasmic fraction (27.39%). A previous study on an Out- mutant (PR54) isolated by Tn*phoA* mutagenesis also revealed a similar cellular distribution of Pel. The fractionation data for this mutant alongside the parent strain is shown in Table 4.3. In this experiment Cel resided in the periplasmic fraction of PR54 whereas Pel was predominantly cytoplasmic. The aberrant localisation of alkaline phosphatase (PhoA) in the parent strain (GS2001) is probably a result of the low levels of this enzyme (<1% wt). The strain (GS2001 was identified as a 'white' colony on NA plates containing XP following Tn*10* mutagenesis. These results, showing a predominantly cytoplasmic accumulation of Pel, contradicted findings by other workers who demonstrated that mutants of the Out- class accumulated Pel predominantly within the periplasm (Andro *et al.*,

Table 4.2. Localisations of extracellular enzymes in an Out- mutant as determined by osmotic shock fractionation

Strain	Enzyme	Enzyme activity (%) in fraction		
		C	P	S/N
HC131	β-gal	82	18	0
(parent)	Bla	9	89	2
	Pel	14	11	75
RJP233	β-gal	85	15	0
(Out-)	Bla	7	90	3
	Pel	64	27	9

Key

C = cytoplasmic fraction
P = periplasmic fraction
S/N = culture supernatant

Table 4.3. Enzyme localisations in GS2001 and Out- mutant PR54 using osmotic shock fractionation

		Enzyme activity (%) in fraction		
Strain	Enzyme	C	P	S/N
GS2001 (parent)	Pel	0	19	81
	Cel	7	22	71
	Prt	10	17	73
	Peh	10	0	90
	PhoA	88	6	6
	β-Gal	60	38	2
	Bla	10	83	7
PR54 (Out-)	Pel	62	38	0
	Cel	12	59	29
	Prt	7	14	79
	Peh	73	16	1
	PhoA	31	68	1
	β-Gal	54	45	1
	Bla	7	82	11

Key

Pel, Cel, Prt and Peh have been defined (Table 1.10.).
PhoA = alkaline phosphatase
β-Gal = β-galactosidase
Bla = β-lactamase
C = cytoplasmic fraction
P = periplasmic fraction
S/N = supernatant fraction

1984; Ji et al., 1987, 1989; Thurn and Chatterjee, 1985; Murata et al., 1990). For this reason other methods of fractionating cells into cytoplasms and periplasms were attempted.

4.3.2. Polymyxin B treatment

Cultures were incubated in the presence of polymyxin B for varying lengths of time using different concentrations of this antibiotic, following the protocol used by Hirst and Holmgren (1987). This method has been used successfully for preparing periplasmic fractions from Ech (Andro et al., 1984).

No significant release of the periplasmic marker enzyme Bla was detected under any conditions (2-10 mg/ml of polymixin B, 5 min, 37°C). However, the quality of release as determined by this method is known to be strain dependent (Hirst, pers. comm.) because some bacteria are resistant to the action of this antibiotic. For this reason this method of fractionation was abandoned.

4.3.3. Fractionation by generating spheroplasts

This procedure was used to remove gently the Gram-negative OM and release the contents of the periplasm without disturbing the IM. The extent of spheroplasting can be conveniently monitored using phase contrast light microscopy. This procedure was adapted from Osborn and Munson (1974) and is described in section 2.17.2.

Enzyme localisations for cultures fractionated using the spheroplasting technique are shown in Table 4.4. The parent strain (HC131) and the 14 EMS Out- mutants were fractionated using this method. Samples were assayed for enzyme activities. In all cases the marker enzymes β-gal and Bla were

Table 4.4. Extracellular enzyme locations in Out- mutants

	Enzyme											
	Pel				Cel				Prt			
	Enzyme activity (%) in cell fraction											
Strain	C	P	S	%wt	C	P	S	%wt	C	P	S	%wt
HC131	2	6	92	100	10	27	63	100	0	0	100	100
RJP249	15	82	3	72	10	90	0	105	0	0	100	37
RJP250	16	80	4	69	10	90	0	105	0	0	100	200
RJP200	22	74	4	70	18	78	4	125	0	0	100	59
RJP220	18	78	4	51	4	98	0	101	0	0	100	73
RJP190	25	74	1	87	14	82	4	164	0	0	100	27
RJP233	14	82	4	34	8	81	11	40	0	0	100	100
RJP251	27	68	5	45	9	91	0	110	0	0	100	40
RJP254	16	79	5	58	18	82	0	159	0	0	100	84
RJP208	19	80	1	115	13	87	0	166	0	0	100	37
RJP253	18	79	3	58	24	76	0	132	0	0	100	94
RJP122	24	74	2	83	57	43	0	122	0	0	100	29
RJP221	11	79	10	29	9	91	0	65	0	0	100	40
RJP211	24	74	2	58	37	63	0	46	0	0	100	49
RJP159	33	66	1	66	26	74	0	167	0	0	100	72
	Enzyme											
	Peh				β-gal				Bla			
	Enzyme acivity in cell fraction											
C	P	S	%wt	C	P	S	%wt	C	P	S	%wt
0	11	89	100	90	10	0	100	8	87	5	100
50	50	0	145	91	9	0	66	6	92	2	98
46	54	0	199	89	11	0	99	7	91	2	126
56	44	0	181	95	5	0	84	8	88	4	110
61	39	0	134	95	5	0	38	6	89	5	86
54	46	0	200	93	7	0	58	12	84	4	107
14	86	0	120	85	15	0	56	5	90	5	139
50	50	0	145	82	18	0	75	10	87	3	110
42	56	2	199	89	11	0	78	4	95	1	142
30	52	18	500	88	11	1	88	6	88	6	109
37	62	1	227	81	18	1	50	12	85	3	101
67	33	0	168	94	6	0	85	14	81	5	134
20	80	0	80	93	7	0	60	3	95	2	128
54	46	0	171	95	5	0	92	9	83	7	115
63	37	0	165	93	7	0	81	16	80	4	113

Key

C = cytoplasmic fraction
P = periplasmic fraction
S = culture supernatant

located predominantly in their appropriate fraction (cytoplasm and periplasm respectively). The normally extracellular enzymes Pel and Cel were present in the supernatant of the parent strain. However, in the Out- mutants, Pel and Cel were predominantly periplasmic. The enzyme Peh was evenly distributed between cytoplasm and periplasm, but this enzyme assay is unreliable. All the Prt activity in all strains detected was present in the supernatant. No intracellular Prt activity was detected.

4.3.4. Discussion of the Out- phenotype

The 14 Out- mutants accumulated Pel and Cel intracellularly (section 4.21.) but secreted Prt normally (section 4.22.). This result strongly suggests, as other workers have also shown (Andro *et al.*, 1984; Ji *et al.*, 1987, 1989; Thurn and Chatterjee, 1985; Murata *et al.*, 1990), that Out- mutants synthesise Pel and Cel normally but are unable to secrete them into the extracellular medium. This pleiotropic phenotype suggests that both Pel and Cel share a common secretory pathway. The fact that Prt secretion is unaltered indicates that this extracellular enzyme is secreted by a different mechanism (section 1.7.4.).

In order to determine the exact site of accumulation of Pel and Cel in Out- mutants, cultures of these mutants were fractionated using osmotic shock, Polymyxin B treatment and also by generating spheroplasts.

After osmotic shock treatment it appeared that Pel was accumulating predominantly within the cytoplasmic fraction. However, there was also a significant proportion (27.5%) of this enzyme activity in the periplasm. This result is in disagreement with those from other workers who showed that this enzyme, along with Cel, accumulated within the periplasm of Out- mutants. One possible reason for this anomaly is that Pel resides in the periplasm of

Out- mutants, but is not released by the osmotic cold shock procedure. Previous osmotic shock studies on another Out- mutant (PR54) showed that Cel was released. A possible explanation for this observation is that the Pel enzymes were accumulating in the periplasm, leading to aggregation when present at an abnormally high concentration within the periplasm. Such aggregations might not be released when Out- mutants are subjected to osmotic cold shock treatment. For this reason two alternative methods of fractionation were employed. The first alternative fractionating procedure utilised polymyxin B but this was not effective for HC131. The other fractionation method employed spheroplasting technology which allowed the disolution of the OM, resulting in the release of the periplasmic contents to the supernatant. Using this method it was demonstrated that the site of accumulation of both Pel and Cel was indeed the periplasm. This result is in agreement with findings by Andro et al. (1984), Ji et al. (1987, 1989), Thurn and Chatterjee (1985) and Murata et al. (1990) who studied Out mutants of Ecc, Eca, and Ech generated using a variety of methods (Tn5, Tn10, Tn10-lacZ, EMS, NTG and mini-Mu mutagenesis). These workers all isolated mutants which accumulated Pel and Cel intracellularly and found the site of accumulation of Pel and Cel to be the periplasm. The findings of this work and of work by others show that Out- mutants accumulate Pel and Cel in the periplasm because they are unable to complete secretion to the extracellular milieu. This explanation assumes that Pel and Cel are secreted by a two-step mechanism involving a periplasmic intermediate (see Figure 1.4.). Work carried out by Andro et al. (1984) demonstrated a build up of Pel in wild-type Ech immediately prior to the onset of its secretion. Work carried out on enterotoxin secretion from Vibrio cholerae demonstrated the presence of a periplasmic intermediate in the secretion of this protein (Hirst

and Holmgren, 1987). However, it cannot be ruled out that Pel and Cel are being artefactually re-routed to the periplasm in Out- mutants of Ecc SCRI193 and that, in wild-type cells, there is no true periplasmic intermediate.

4.4. Phage sensitivity in Out- mutants

4.4.1. Introduction

Work carried out in this laboratory by I. Toth and others has led to the isolation of several bacteriophages which infect individual Erwinia strains used in this laboratory. Using enrichment techniques it was possible to isolate, from sewage, four bacteriophages which infect Ecc SCRI193. One of these bacteriophages, ØKP, was later shown to be a generalised transducing phage for this strain and also Ecc strain ATTC39048. There was a possibility that proteins involved in Pel and Cel secretion across the OM were located in the OM of Ecc. It has already been mentioned that some of the proteins of the general export pathway are located in the IM and are involved in translocating proteins across this membrane. Also, it is generally known that the E. coli bacteriophage, enters the cell by initially adsorbing to the LamB protein, an OM protein involved in the transport of maltose. Furthermore, some mutations in the LamB protein are known to be pleitropically deficient in the ability to transport maltose and adsorb phage λ (Charbit et al., 1988). The ability of these four Ecc bacteriophages to infect the secretion mutants was tested.

4.4.2. Bacteriophage spot tests and efficiency of plating (eop) assays

Spot tests using each of the four independently isolated Ecc

bacteriophages were performed on bacterial lawns of the Out- mutants. This was carried out by spotting 10 µl of a lysate of each bacteriophage onto a bacterial lawn of each mutant as described in section 2.12. This crude test was carried out using low titre bacteriophage stocks (1 x 10^4 pfu/ml).

The results of this experiment are given in Table 4.5. Two of the bacteriophages (ØKP and Ø575) infected the parent strain HC131 and all Out- mutants. However, the other two bacteriophages (Ø565 and ØD-2) produced no zones of lysis on the Out- mutant RJP190 but grew on the other 13 Out- mutants and the parent strain HC131.

A photograph showing the results from an efficiency of plating (eop) assay is given in Figure 4.1. Two plates with bacterial lawns of two strains (HC131 [parent] and RJP190 [Out-] respectively) are shown. Bacteriophage stocks were serially diluted and spotted onto the lawns as illustrated. Bacteriophages Ø575 and ØKP grew on HC131 and RJP190 to the same extent. However, ØD-2 and Ø565, which grew on HC131 did not grow on RJP190.

High titre lysates of ØD-2 and Ø565 were then prepared (section 2.8.). These lysates were serially diluted and used to infect lawns of HC131 and RJP190 in order to determine their efficiency of plating on these strains. The results from this experiment are shown in Table 4.6. Bacteriophage ØD-2 infected to a titre of 4.2 x 10^8 pfu/ml on HC131 but only to a titre of 4.2 x 10^6 pfu/ml on RJP190. This means that ØD-2 infects RJP190 99% less efficiently than it does HC131. Bacteriophage Ø565 infected to a titre of 5.0 x 10^6 on HC131 compared to 1.0 x 10^3 on RJP190. This represents a 99% lowering of the efficiency of plaque formation of Ø565 on RJP190 compared to growth on the parent strain HC131. Also, the plaque morphologies of both ØD-2 and Ø565 were altered on RJP190 when compared to HC131. The plaques were smaller and more turbid in appearance suggesting that the

Table 4.5. Sensitivity of Out- mutants to four Ecc bacteriophages using spot tests

Strain	Bacteriophage			
	ØKP	Ø565	ØD-2	Ø575
HC131	+	+	+	+
RJP253	+	+	+	+
RJP190	+	-	-	+
RJP233	+	+	+	+
RJP200	+	+	+	+
RJP251	+	+	+	+
RJP254	+	+	+	+
RJP250	+	+	+	+
RJP211	+	+	+	+
RJP249	+	+	+	+
RJP221	+	+	+	+
RJP122	+	+	+	+
RJP220	+	+	+	+
RJP159	+	+	+	+
RJP200	+	+	+	+

Key

+ Bacteriophage plates on strain

- Bacteriophage does not plate on strain

Figure 4.1. Efficiency of plating of bacteriophages on the Out- mutant RJP190

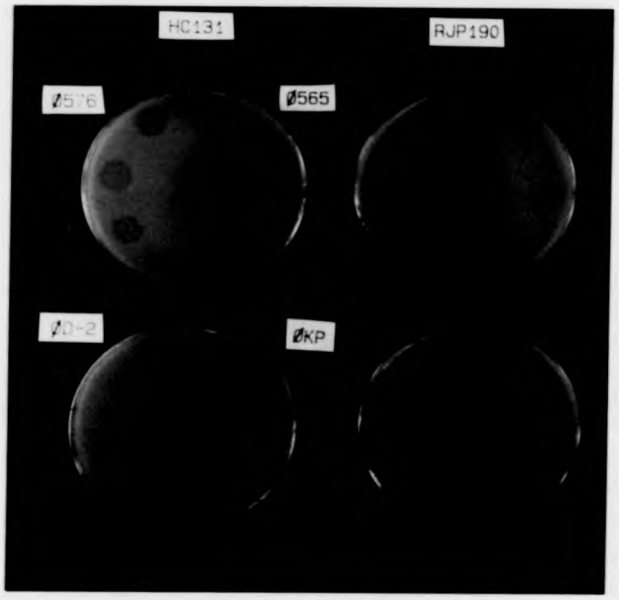


Figure legend

Two bacterial lawns of each HC131 (parent) and RJP190, were challenged with serial dilutions of four different Ecc bacteriophages. Four dilutions (10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}) of each bacteriophage were spotted onto each half of the plate as indicated. The lowest dilution was placed at the top half of each plate and the highest dilution at the bottom half as shown.

Table 4.6. Efficiency of plating (eop) of *Ecc* bacteriopages on Out- mutants

Strain	Eop of *Ecc* bacteriophage (pfu/ml lysate)	
	ϕD-2	ϕ565
HC131	4×10^8	5×10^6
RJP190	4×10^6	1×10^3
AC400	4×10^8	4×10^6
PR54	3×10^8	4×10^6

Figure legend

Bacteriophages ϕD-2 and ϕ565 were re-prepared to produce high titre lysates, serially diluted and used to infect the parent strain (HC131) and the above Out- mutants. Out- mutants AC4000 (HC131::Tn*10*) and PR54 (GS2001::Tn*5*) were also tested as they were not tested in the previous experiment (Table 4.5.).

phages had difficulty in propagating on RJP190. These results reinforce the findings from the preliminary phage spot test in that Out- mutant RJP190 is altered in its susceptibility to infection by ØD-2 and Ø565. However, due to the high titres of each bacteriophage in this experiment it was shown that they were both still able to infect RJP190 but at a lower frequency.

4.4.3. Phage adsorption assays

This experiment was performed to investigate further the lowered plating efficiency of ØD-2 on Out- mutant RJP190.

Phage adsorption assays were carried out as described in section 2.12. using ØD-2. This particular bacteriophage was used because it produced large and clear plaques which were easy to identify and quantify. The results of a phage adsorption assay using mutants RJP190 and RJP253 are given in Table 4.7. The controls used were Ecc HC131 and E. coli DH1 (positive and negative respectively). Out- mutant RJP253 was also used as a control. The original bacteriophage titre of ØD-2 was 6.0^{10} pfu/ml.

The pfu/ml at t=0 min and t=25 min represent the titre of bacteriophage ØD-2 in the bacterial supernatant at those two times. The change in bacteriophage titre over that time is assumed to result from the bacteriophage adsorbing to the surface of the bacterial cell. After adsorption, the number of 'free' bacteriophage particles remaining in the bacterial supernatant (unadsorbed) after pelleting the bacteria is determined by titrating this supernatant on the sensitive parent strain HC131.

The results of this experiment are shown in Table 4.7. Over the time period allowing for bacteriophage adsorption (20 min), 83% of the infective bacteriophage particles were removed by the parent strain HC131. Mutant RJP253 also had a similar effect as HC131 and removed 73% of the

Table 4.7. Efficiency of plating (eop) of bacteriophage ϕD-2 post-adsorption

Strain	Eop of bacteriophage ϕD-2			
	ϕD-2 t=0 min pfu/ml	ϕD-2 t=25 min pfu/ml	ϕD-2 pfu/ml (adsorbed)	Adsorption (%)
HC131	6.0×10^{10}	1.0×10^{10}	5.0×10^{10}	83
DH1	6.0×10^{10}	7.0×10^{10}	0	0
RJP190	6.0×10^{10}	7.0×10^{10}	0	0
RJP253	6.0×10^{10}	2.0×10^{10}	4.0×10^{10}	67

Legend to Table 4.7.

The bacterial strains used in this experiment are shown in the first column. The next two columns show the change in titre (pfu/ml) of bacteriophage ϕD-2 between two time points, t=0 min (pre-adsorption) and t=25 min (post-adsorption). The difference between these two values (column four) gives the number of bacteriophages removed from the bacterial supernatant by a particular strain. The final column shows the level of adsorption. This is expressed as the proportion of bacteriophage removed from the original lysate.

bacteriophage from the supernatant. The negative control E. coli DH1 did not adsorb Ecc bacteriophage ϕD-2 within the time period allowed. Similarly, mutant RJP190 did not adsorb ϕD-2 within the allowed time period.

4.4.4. Discussion of the phage resistant phenotype

The pleiotropic phenotype [ϕ^r, Out-] has not been previously reported. Only one (RJP190) of the Out mutants, which were all similar with respect to their extracellular enzyme phenotype, showed resistance to the Ecc bacteriophages ϕD-2 and ϕ565. This implies that more than one protein is involved in Pel and Cel secretion. Mutations in any of these proteins might give rise to the Out- phenotype. It is interesting to speculate that the mutation in RJP190 is in a gene encoding a protein located in the OM, which is involved in Pel and Cel secretion, and also acts as the receptor for ϕD-2 and ϕ565 adsorption. An alternative explanation might be that this mutant is unable to correctly assemble another protein or molecule which acts as the receptor for these two bacteriophages. In theory it should be possible to isolate Ecc mutants resistant to infection by ϕD-2 and ϕ565. Some of these mutants might also be expected to be Out-. This might provide a method for enriching Out- mutants.

4.5. Pathogenicity assays

Pathogenicity tests were carried out on potato tubers using the 14 Out- mutants following the method described in section 2.11. A negative control (E. coli LE392) and a positive control (Ecc HC131) were also tested using this assay. Strains were inoculated in duplicate and the average 'rot' (diameter of area of rot) for each strain is given in Table 4.8. All

Table 4.8. Pathogenicity assays using potato tubers

Strain	Level of tissue maceration			
	Inoculum 1 (mm)	Inoculum 2 (mm)	Rot (average) (mm)	Degree of rot (% wt)
HC131	15	18	16.5	100
LE392	0	0	0	0
RJP200	16	14	15	91
RJP249	9	16	12.5	76
RJP253	8	0	4	24
RJP250	18	19	18.5	112
RJP221	0	6	3	18
RJP251	0	0	0	0
RJP254	0	4	2	12
RJP190	8	15	11.5	70
RJP159	8	3	5.5	33
RJP233	8	7	7.5	45
RJP208	11	3	7	42
RJP122	15	3	9	55
RJP211	5	7	6	36
RJP220	6	3	4.5	27

Legend

The strains used in this experiment were HC131 (parent strain- positive control), LE392 (E.coli - negative control) and the 14 EMS derived Out- mutants. Further details are provided in the text.

inoculations were performed in duplicate using the same potato tuber for each mutant tested.

4.5.1. Discussion

The data in Table 4.8. suggests that there was a lowered level of 'rot' caused by the Out- mutants compared with the parent HC131. The negative control *E. coli* LE392 did not cause any noticeable rot. This result confirms that *E. coli* cannot rot potato tubers and demonstrates that secondary infection by pathogens capable of causing rot did not occur. The varying degrees of rot caused by different Out- mutants is interesting. However, this experiment is crude and difficult to control. For example, some tubers might be more susceptible to soft rot than others. These results are regarded as preliminary findings and might indicate that Out- mutants have a lowered ability to macerate potato tubers. A similar experiment carried out by other workers showed that *Ecc* 71 Out- mutants caused less soft rot than wild-type cells (Murata *et al.*, 1990). These results suggest that the correct localisation of Pel and Cel by *Ecc* is an important factor in the aetiology of potato soft rot.

4.6. Summary of the Out- phenotype

Out- mutants accumulate Pel and Cel intracellularly and the exact site of accumulation has been identified as the periplasm. It has also been demonstrated that Out- mutants appear to be impaired in the ability to 'rot' potato tubers under laboratory conditions. Furthermore, one Out- mutant, RJP190, was shown to have a reduced sensitivity to two *Ecc* bacteriophages, ϕD-2 and ϕ565. Results from bacteriophage adsorption experiments suggested

that the reason for phage resistance was due to the decreased adsorption of ØD-2 and Ø565 to the cell surface of RJP190.

CHAPTER 5

THE GENETIC ANALYSIS OF Out- MUTANTS

5.1. Introduction

In the previous chapter the physiological characterisation of 14 independently isolated Out- mutants was described. This chapter will focus on the genetic analysis of the Out- mutants. As mentioned in section 1.9.12., a derivative of Ecc SCRI193 has been constructed which is sensitive to bacteriophage λ infection. This strain (HC131) harbours plasmid pHCP2 which encodes the LamB protein. HC131 is susceptible to bacteriophage λ adsorption and subsequent DNA injection. However, the coliphage λ cannot replicate in Ecc, but can be used as a suicide vector for the delivery of transposons and cosmids (see Figure 1.12.). The Out- mutants are HC131 derivatives and can be transduced with cosmids using λ. By introducing a fully representative cosmid library into Ecc mutants, it should be possible to directly complement these mutants and restore the wild-type phenotype. This is a powerful technique and enables genes of interest to be cloned onto large DNA fragments. This approach was used in an attempt to complement Out- mutants and isolate the out+ gene(s).

5.2. In vivo packaging of a cosmid library into λ

The cosmid pHC79 (see Appendix I) was used to construct a genomic library of Ecc using Ecc chromosomal DNA which had been partially digested with EcoRI. This Ecc cosmid library was constructed in this laboratory by M. Gibson (pers. comm.). This library has been used previously to isolate pel+ genes (Plastow et al., 1986), cel+ genes (V. Cooper, pers. comm.), and rec+ genes (M. Gibson, pers. comm.) from Ecc.

The cosmid library was packaged in vivo into λ heads using λcI857. An E. coli DH1 culture containing a heterogeneous Ecc library was used to

prepare the λ cosmid lysate. This method is described in section 2.13.1. A λ lysate was made and upon infection of Ecc HC131 (section 2.13.2) 200 to 1000 Tc^r transductants per ml of lysate were obtained.

5.3. Complementation of Out- mutants using a λ delivered Ecc cosmid library

5.3.1. Screening for complementation

The Ecc cosmid library was introduced into the λ sensitive Out- mutants as described in section 2.13.2. Transductants were selected using Tc^r, a pHC79 encoded resistance marker. The introduction of cosmid pHC79 was concomitant with the loss of the incompatible LamB encoding plasmid, pHCP2. Transductants were screened for the return of an Out+ phenotype on the extracellular enzyme detection plates (section 2.9.). At least 200 Tc^r transductants were screened for each mutant.

5.3.1.1. Results

Twelve of the 14 Out- mutants were complemented using the λ cosmid lysate as determined by qualitative plate assays. Complemented Out- mutants were identified by the presence of halos surrounding colonies on the Pel and Cel enzyme detection plates. A list of the Out- mutants restored to Out+ by transduction with the Ecc cosmid library is presented in Table 5.1. The frequencies of appearance of Out+ colonies for each transduced Out- mutant are also given. Figure 5.1. shows photographic evidence of Out- mutants and the same mutants harbouring complementing cosmids. The halo surrounding the centre colony on the Pel, Cel and Prt detection plates signifies the production of these enzymes by the parent strain HC131. Arranged around the parent colonies are three Out- mutants (1 = RJP190, 2 = RJP253, 3 = RJP159)

Table 5.1. Cosmid complementation of Out- mutants

Mutant	No. transductants screened	No. of Out+ transductants	Frequency (%)
RJP200	700	0	0.0
RJP254	200	4	2.0
RJP249	200	1	0.5
RJP190	400	4	1.0
RJP122	600	8	1.3
RJP221	200	3	1.5
RJP220	200	1	0.5
RJP251	200	0	0.0
RJP250	200	1	0.5
RJP208	200	2[a]	1.0
RJP159	200	4	2.0
RJP233	200	1	0.5
RJP211	200	3	1.5

a These transductants were not fully restored to Out+ but appeared to be partially complemented.

Figure 5.1. Complementation of Out- mutants shown by plate assays

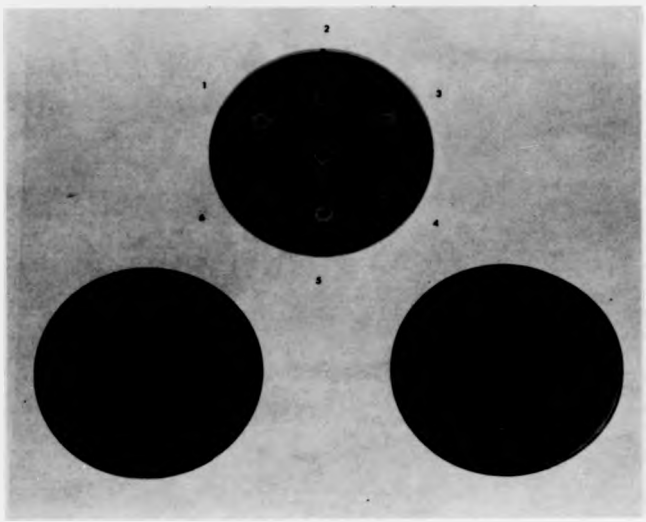


Figure legend

brown plate = protease detection
dark blue plate = cellulase detection
light blue plate = pectinase detection

Colony	Strain name	Phenotype
centre	HC131	wild type (Out+)
1	RJP190	Out-
2	RJP253	Out-
3	RJP159	Out-
4	RJP190(cHIL190)	Out+
5	RJP253(cHIL253)	Out+
6	RJP159(cHIL159)	Out+

and the same three mutants harbouring complementing cosmids (4 = RJP159, 5= RJP253, 6 = RJP190). The Out- mutants (1, 2 and 3) are easily identified by the absence of halos surrounding these colonies. Complemented Out- mutants (4, 5 and 6) have a wild-type (Out+) phenotype on Pel and Cel plates. Extracellular protease is produced by the three Out- mutants in this photograph. The production of Prt by colonies 2 and 3 (RJP253 and RJP159) might be reduced in comparison to the parent strain. This effect is unaltered for the same two mutants restored to Out+ by the complementing cosmids.

5.3.2. Confirmation that the Out+ phenotype is restored in Out- mutants harbouring complementing cosmids by using quantitative enzyme assays

Out- mutants harbouring putative complementing cosmids, as determined by plate assays, were grown in extracellular enzyme inducing medium (PM) and fractionated into whole cell extracts and supernatants as described in section 2.16. These fractions were assayed for Pel (section 2.10.3.) and Cel (section 2.10.6.1.) activity. The intracellular enzyme β-lactamase (Bla) was also assayed (section 2.10.2.) to ensure that the complemented mutants were not exhibiting a general leaky phenotype.

5.3.2.1. Results

The results from this experiment are shown in Table 5.2. Enzyme activities in each fraction (cell sonicate or supernatant) are expressed as a percentage of the total activity of the combined fractions. At the time of this experiment RJP233, RJP251, RJP200 and RJP211 had not been complemented. RJP208 was not assayed because this Out- mutant had not been fully restored to Out+ by a putative complementing cosmid. In all samples, Bla was predominantly intracellular, except in RJP249 and RJP250

Table 5.2. Enzyme localisations in Out- mutants and Out- mutants harbouring complementing cosmids

	Enzyme								
	Pel			Cel			Bla		
	Enzyme activity (%) in cell fraction								
Strain	SO	SN	%wt	SO	SN	%wt	SO	SN	%wt
HC131	8	92	100	22	78	100	97	3	100
RJP254	97	3	275	90	10	37	96	4	93
RJP254(c)	6	94	57	7	93	84	97	3	119
RJP249	99	1	109	100	0	28	98	2	84
RJP249(c)	30	70	12	7	93	70	49	51	19
RJP190	97	3	154	100	0	22	98	2	98
RJP190(c)	21	79	87	22	78	100	94	6	88
RJP122	99	1	173	100	0	33	98	2	91
RJP122(c)	43	57	41	11	89	74	95	5	104
RJP220	100	0	125	100	0	51	99	1	119
RJP220(c)	10	90	37	7	93	84	95	5	49
RJP250	99	1	201	100	0	50	97	3	116
RJP250(c)	80	20	30	63	33	13	63	37	60
RJP159	99	1	144	100	0	25	97	3	95
RJP159(c)	2	98	44	10	90	37	98	2	146
RJP221	98	2	44	100	0	9	99	1	95
RJP221(c)	23	77	18	50	50	10	96	4	91
RJP253	99	1	90	100	0	22	98	2	84
RJP253(c)	11	89	103	7	93	84	98	2	125

Key

SO = cell sonicate fraction

SN = cell supernatant fraction

(c)= mutant harbouring 'putative' complementing cosmid

%wt= total enzyme activity compared to the wild type (parent) strain HC131

harbouring putative complementing cosmids. Significant amounts of Bla (51% and 37% respectively) were present in the cell supernatant of these two strains. The localisation of Pel and Cel for the wild-type strain HC131 was as expected, with both enzymes residing predominantly in the cell supernatant. The Out- mutants were all found to accumulate Cel and Pel in the cell sonicate fraction with only small amounts of these enzymes reaching the cell supernatant. However, both Pel and Cel were located predominantly in the supernatant when Out- mutants carrying complementing cosmids were assayed. Indeed, the cellular distribution of these two enzymes in complemented Out- mutants had been restored to that of the wild-type strain HC131.

5.3.3. Discussion

The Out+ phenotype was restored in 11 of the 14 Ecc Out- mutants using an Ecc cosmid library as determined by plate assays. It was not possible to restore the Out+ phenotype in three of the 14 Out- mutants, these being RJP251, RJP200 and RJP208. However, two RJP208 transductant colonies were isolated which appeared to show an intermediate level of Pel and Cel production when compared to Out- mutants and the wild-type strain (HC131) as determined by plate tests. These colonies were saved because of this curious phenotype.

The frequency of Out+ colonies found upon screening the Tc^r transductants was approximately 1% but varied for each Out- mutant. Cosmids are expected to package approximately 40 kb of DNA (1% of total genome) and in theory any random chromosomal marker would be expected to be present at a frequency of 1%. The results of this study are in agreement with this expected frequency.

Nine putatively complemented mutants were subjected to cellular fractionation followed by quantitative enzyme assays. At the time of these investigations RJP233 and RJP211 had not been complemented but were complemented later as determined by plate assays. Cosmid complementation was confirmed for seven of the nine putatively complemented mutants by assaying for extracellular enzymes in cell sonicate and cell supernatant fractions using quantitative enzyme assays. The complemented Out- mutants displayed a wild-type distribution of Pel and Cel and retained Bla activity within the cell fraction. In general, the mutants harbouring complementing plasmids displayed reduced levels of total enzyme activity for Pel and Cel when compared to the wild-type strain HC131. The reason for this is not known but might be the result of cellular stress caused by the multiple copy cosmid expressing abnormally high levels of 'exported' Out proteins. Two of the putatively complemented mutants, RJP250(cHIL250) and RJP249(cHIL249), had high levels of Bla in the culture supernatant, indicating that this pseudo-complementation was the result of non-specific leakage of cellular contents into the supernatant. Upon closer examination, these putatively complemented mutants did not show wild-type levels of Pel and Cel production when assayed using plate tests.

The cosmids identified on the basis of being able to restore the Out+ phenotype in Out- mutants, without causing cell leakiness, were assumed to carry out+ genes. Such cosmids will be referred to as out+ cosmids.

5.4. Classical genetic analysis of cosmids which restore the Out+ phenotype in Out- mutants

The aim of this experiment was to identify subclasses within the Out- class of mutants. The approach taken was to introduce each independently

isolated out+ cosmid into each of the other 13 Out- mutants. If there was more than one out gene and these out genes were separated by a length of DNA beyond the limits allowed for packaging by λ, it should be possible to identify distinct complementation groups. The ten cosmids from the complemented Out- mutants (including the cosmid which partially restored the Out+ phenotype in RJP208) were isolated from these strains using modified alkaline lysis procedure for Erwinia described in section 2.18.1. In order to simplify analysis and enable its repackaging into λ, cosmid DNA from Ecc Out- mutants was used to transform the recA- strain of E. coli DHI (section 2.26.). Cosmid lysates were then prepared by propagating λc1857 on E. coli strains harbouring out+ containing cosmids, as described in section 2.13.1. The names of the out+ cosmids were derived from the Out- mutant that they were originally shown to complement. For example, cHIL190 was isolated from a complemented (Out+) colony of Out- mutant RJP190.

All the Out- mutants were infected with each cosmid lysate and, as before (section 5.3.1.), transductants were selected on NA plates containing Tc. Five Tc^r colonies from each separate transduction were then screened on the Pel and Cel extracellular enzyme detection plates.

5.4.1. Results

The results of this experiment are shown in Table 5.3. and are summarised in Figure 5.2. These results show that the complementing cosmids fall into three separate classes which complement three distinct groups of mutants. Interestingly, three mutants, RJP200, RJP251 and RJP249, which had not been complemented when transduced with the cosmid library, were complemented strongly with cosmids which complemented other Out- mutants. At this stage these complementation groups will be called group A,

Table 5.3. Cross-complementation pattern of Out- mutants when transduced with cosmids encoding Out functions

	Cosmid lysate(cHIL)									
Mutant	122	190	211	233	220	221	254	159	253	208
RJP250	-	-	-	-	-	-	-	-	-	(+)
RJP208	-	-	-	-	-	-	-	-	-	(+)
RJP122	+	+	+	+	-	-	-	-	-	(+)
RJP190	+	+	+	+	-	-	-	-	-	(+)
RJP211	+	+	+	+	-	-	-	-	-	(+)
RJP200	+	+	+	+	-	-	-	-	-	(+)
RJP233	+	+	+	+	-	-	-	-	-	(+)
RJP220	+	+	+	+	+	+	+	+	+	(+)
RJP221	+	+	+	+	+	+	+	+	+	(+)
RJP254	+	+	+	+	+	+	+	+	+	(+)
RJP159	-	-	-	-	-	+	+	+	+	(+)
RJP249	-	-	-	-	-	+	+	+	+	(+)
RJP253	-	-	-	-	-	+	+	+	+	(+)
RJP251	-	-	-	-	-	+	+	+	+	(+)

Key

\+ = complemented mutant
\- = non complemented mutant
(+) = partially complemented mutant

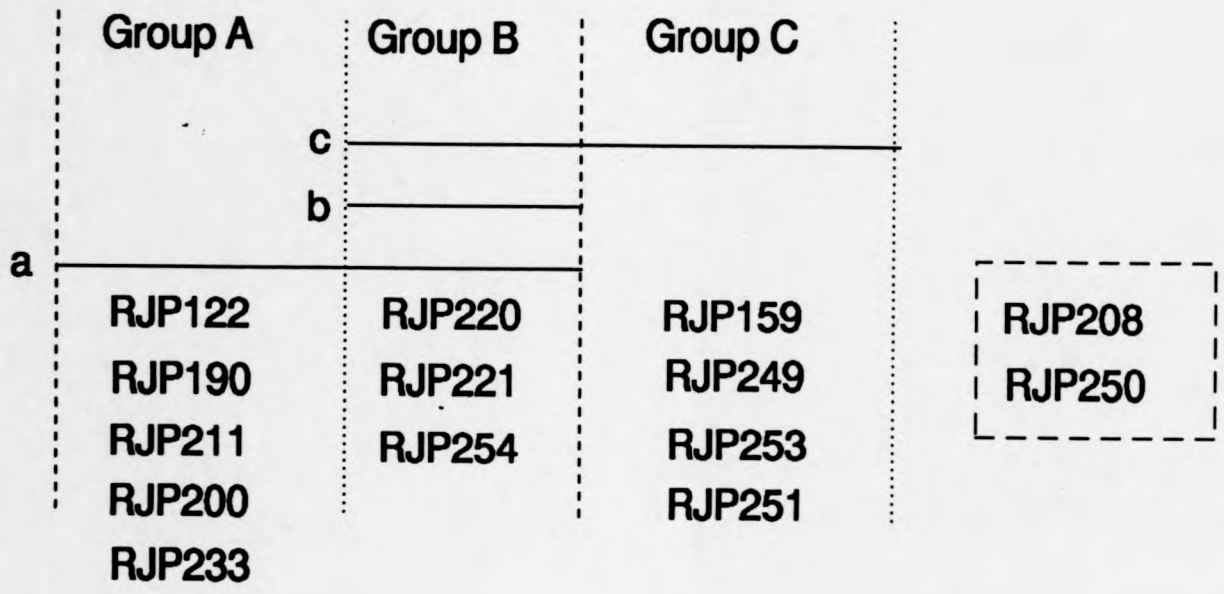
Legend

Five colonies from each of the 144 seperate transductions were screened on extracellular detection plates for the production of Pel and Cel activity.

Legend for Figure 5.2.

The three complementation groups identified have been classified as group A, group B and Group C. The Out- mutants belonging to these three groups are listed beneath their respective group. The limits of each complement group is represented by a dotted lines. The DNA carried by complementing cosmids is shown by lines with lower case letters which define groups of cosmids. Type a cosmids (cHIL122, cHIL90, cHIL211 and cHIL233) can complement Group A and Group B Out- mutants. A type b cosmid (cHIL220) was able to complement group B Out- mutants only. Type c cosmids (cHIL221, cHIL254, cHIL159 and cHIL253) were able to complement Group B and Group C Out- mutants. Out- mutants RJP208 and RJP250 were never properly complemented and might define another complementation group.

Figure 5.2 Complementation groups of Out- mutants



group B and group C.

Group A consisted of Out- mutants RJP122, RJP190, RJP211, RJP200 and RJP233. This group of mutants was complemented by the following cosmids: cHIL122, cHIL190 and cHIL233 (group a cosmids).

Group B consisted of Out- mutants RJP220, RJP221 and RJP254. Mutants in this class were complemented fully by cHIL220 and all the cosmids which complemented Out- mutant classes A and B. Cosmid cHIL220 (group b cosmid) only complemented group B Out- mutants, thus enabling the identification of this group.

Group C consisted of Out- mutants RJP159, RJP249, RJP253 and RJP251. Mutants in this class were complemented by cosmids cHIL221, cHIL254, cHIL159 and cHIL253 (group c cosmids).

Out- mutants RJP251 and RJP208 were never fully complemented by any cosmid. However, they were both partially complemented by cHIL208. Cosmid cHIL208 also partially complemented all other mutants, but the halos on enzyme detection plates were always of intermediate size for both Pel and Cel. The cosmids which were originally thought to complement RJP249 and RJP250 (but in fact caused cell leakiness, see Table 5.2) were also introduced into all the Out- mutants but did not complement any.

5.4.2. Discussion

Using a classical genetic approach of direct complementation, it was possible to assign 12 of the Out- mutants to three distinct complementation groups (A, B and C). As it was not possible to fully complement two other Out- mutants (RJP251 and RJP208), there existed a possibility that at least one more *out* locus existed. The complementation pattern revealed linkage between these groups. Cosmids able to complement group A mutants group a

cosmids) also complemented group B mutants. Cosmids complementing group C mutants (group c cosmids) also complemented group B mutants. However, one cosmid, cHIL220 (group b cosmid), only complemented the mutants of group B. A diagram illustrating the possible organisation of *out* loci is given in Figure 5.2. This diagram indicates that some of the cosmids might have been expected to share common fragments of DNA, an hypothesis which was investigated and is discussed in the following section.

5.5. Physical characterisation of out+ cosmids - restriction mapping analysis

5.5.1. Introduction

The next step was to analyse the DNA of *out*+ cosmids. The cosmid library was constructed by 'shotgun cloning' partially *Eco*RI digested *Ecc* chromosomal DNA into the unique *Eco*RI site of the cosmid vector pHC79 (M. Gibson, pers. comm.). It was hoped that by digesting the *out*+ cosmids with *Eco*RI and other restriction enzymes, the insert profile for each cosmid, the size of the DNA insert, and also the presence of common restriction fragments within the different *out*+ cosmids might be determined.

Cosmid DNA from the *out*+ cosmids was prepared from the appropriate *E. coli* DHI strain carrying *out*+ cosmids cHIL159, cHIL122, cHIL253, cHIL190 and cHIL220. The DNA from *E. coli* DH1 carrying cosmid cHIL208 was also prepared. All cosmids were prepared using the 'midi-preparation' procedure described in section 2.18.3. The DNA was cut with various restriction endonucleases (section 2.19.) and analysed by agarose gel electrophoresis (section 2.20.).

5.5.2. Results

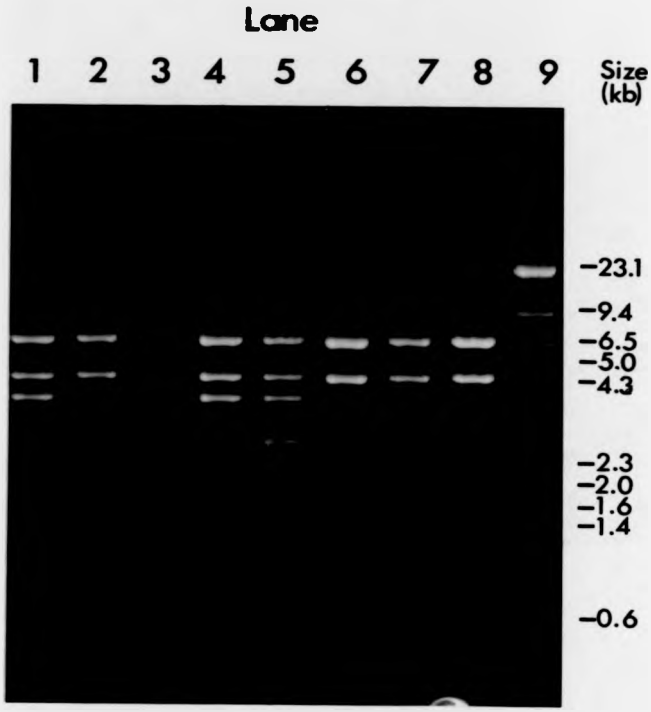
5.5.2.1. Preliminary mapping with EcoRI

An agarose gel showing the restriction pattern obtained after digestion with EcoRI is shown in Figure 5.3. When digested with EcoRI, the banding patterns of each of the three groups of cosmids revealed the presence of some common DNA fragments. The vector band (6.5 kb) can clearly be seen in all tracks as can another common band migrating at 4.5 kb. Cosmid cHIL220 (group b) has only this 4.5 kb fragment, whereas group (a) and group (c) have this fragment and other bands unique to them. These are 0.8 kb and 3.7 kb (group a) and 1.6 kb and 1.4 kb (group b). Cosmid cHIL208 contained all the bands common to the three other out+ cosmid groups (a, b and c) and also extra unique bands. This suggested that cHIL208 spanned the entire region of DNA covered by cosmids cHIL159, cHIL220 and cHIL253. These results allowed the construction of a schematic representation of the out complementation groups and their positions on a rudimentary genetic map. These findings are shown in Figure 5.4.

5.5.2.2. Further restriction mapping of cosmid cHIL159

The restriction mapping of the three groups of cosmids which spanned 12 kb of DNA was carried out by D. Whitcombe (pers. comm.). The restriction map obtained is shown in Figure 5.5. The restriction map of the entire region 12 kb region (carried by group a, b and c cosmids but not the additional DNA carried by cHIL208) has now been confirmed by DNA sequence analysis.

Figure 5.3. Digestion of out+ cosmids with EcoRI



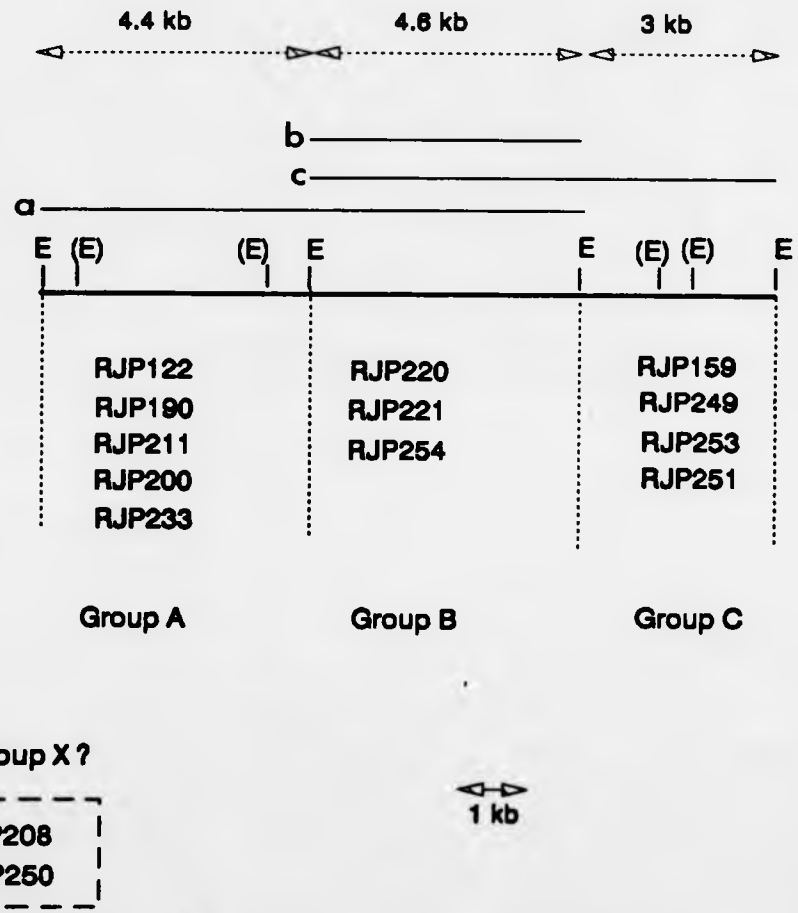
Key

Lane	Restriction digest
1	cHIL122 EcoRI
2	cHIL253 EcoRI
3	cHIL159 EcoRI
4	cHIL190 EcoRI
5	cHIL208 EcoRI
6	cHIL220 EcoRI
7	cHIL253 EcoRI
8	cHIL220 EcoRI
9	Mw markers

Legend to Figure 5.4.

Figure 5.4 shows a rudimentary EcoRI restriction map of the out+ DNA carried by out+ cosmids. The EcoRI restriction enzyme sites in brackets (E) represent possible recognition sites for this restriction enzyme. The three out+ cosmid groups (group a, group b and group c) which carry DNA spanning this 12 kb region are also shown in this diagram.

Figure 5.4. Restriction mapping cosmids using *Eco*RI

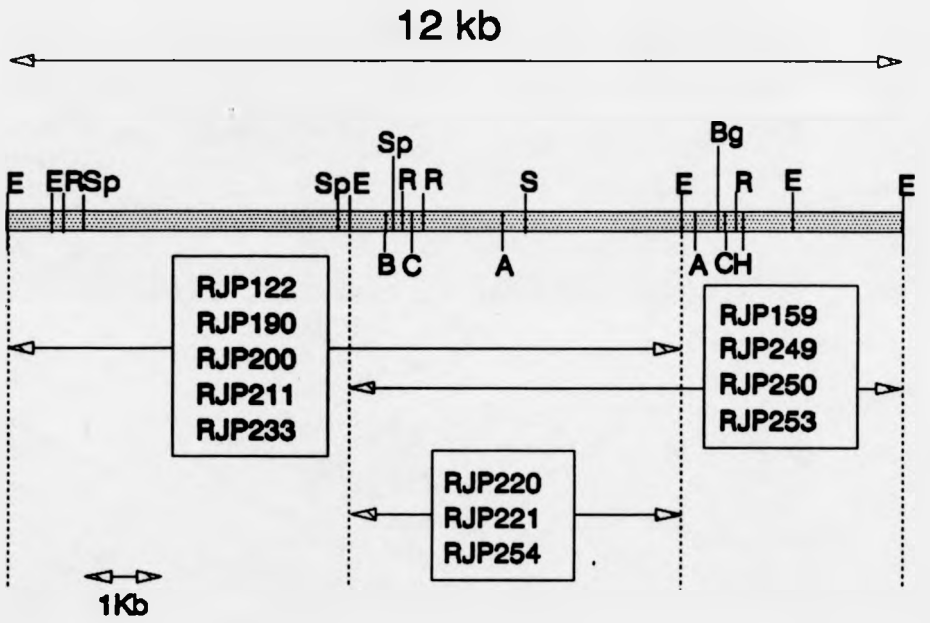


Legend to Figure 5.5.

Figure 5.5. shows a restriction map of DNA containing the out gene cluster. Out- mutants belonging to the three complentation groups (see Figure 5.4) are boxed. The abbreviations for restriction endonucleases are listed below.

A	AvaI
B	BamHI
Bg	BglII
C	ClaI
E	EcoRI
H	HindIII
R	EcoRV
S	SalI
Sp	SphI

Figure 5.5. Restriction map of the *out* gene cluster



5.5.2.3. Discussion

The initial digestion of cosmids from three different groups (a, b and c cosmids) with EcoRI provided some interesting and surprising results. As was anticipated, the three different cosmids did indeed share common fragments of DNA. Cosmid cHIL220, which complemented the group B mutants only, contained a single EcoRI fragment of only 4.5 kb. This fragment was also present in all the other cosmids, thus explaining why they were all (except cHIL208) able to complement the group B mutants. Similarly, the unique fragments of DNA belonging to the other out+ cosmids (groups a and c) were assumed to encode the complementing Out functions for the corresponding Out- mutants. Using this reasoning it was possible to assign certain Out- mutants to three particular regions on this piece of DNA. It follows that the other two EMS mutants (RJP208 and RJP251) might be expected to lie outside this region.

The size of the DNA fragment covered by the three different cosmids (cHIL208 apart) was predicted to be 12 kb. This fragment of DNA fully complemented 12 of the EMS generated Out- mutants and a TnphoA Out- mutant, PR33 (D. Whitcombe, pers. comm.). Another TnphoA derived Out- mutant, PR54, was not complemented by any cosmids but was partially complemented by cHIL208 (D. Whitcombe pers. comm.).

Some intriguing questions arise from this work so far. Why does cHIL208 not 'fully' complement any of the Out- mutants when it clearly contains the same DNA fragments? Furthermore, how can such small fragments of DNA (4.5 kb - 9 kb plus vector [8.5 kb]) be packaged into bacteriophage λ which normally packages 48 kb of DNA (Sternberg and Weisberg, 1975). It was demonstrated by S. Wharam (pers. comm.) that the orientation of the DNA insert in cHIL208 was opposite to that of the all the

other cosmid inserts. This might suggest that in the complementing cosmids, the cloned out+ genes are being expressed from promoters within the vector DNA. The second question involving the cloning of such small fragments of DNA is more difficult to explain. Another research group within this department have also come across such anomalies (C. Murrell, pers. comm.). A possible explanation might be that higher molecular weight concatemers are being packaged which give the appearance of lower weight molecules upon digestion.

5.8. Subcloning out genes which complement group C mutants

5.8.1. Using deletions to clone out genes

5.8.1.1. Introduction

This work was carried out using information from the restriction map. A degree of subcloning had already occurred fortuitously as a result of the aberrant packaging of small cosmids by bacteriophage λ. This led to the identification of four complementation groups (including at least one group to which RJP251 and RJP208 must belong). The next step was an attempt to further subdivide the C group of mutants with the aim of defining more complementation groups.

The cosmid pHC79 has several unique restriction sites which cut within the Tc^r gene (see Appendix I). The presence of another restriction site for the same restriction enzyme within the DNA insert would enable the deletion of some of the Tc^r gene (thus resulting in its inactivation) and some of the DNA insert (which might lead to the loss of out+ genes). Using ligation conditions to enhance the re-ligation of the remains of the original plasmid

without the re-insertion of the 'deleted' fragment (section 2.25.1.), it was possible to generate deletant derivatives of the original cosmid. The theory of this approach is illustrated in Figure 5.6.

Putative cosmid deletants were introduced into the mobilising strain of E. coli GJ342 by transformation (section 2.26.) and mobilised into Ecc Out- mutants by selecting for Ap^r. Out- mutants had previously been cured of pBR322 based plasmids (section 2.14.). Transconjugants were selected on MM containing sucrose as the carbon source (to counter-select the auxotrophic E. coli mobilising strain GJ342) and Ap (to select for Ecc transconjugants). Ecc Ap^r transconjugants were then screened on enzyme detection plates (to identify Out- mutants restored to Out+) and NA plates containing Tc (to determine whether a deletion had occurred).

Colonies with an Out+ Ap^r Tc^s phenotype were purified by re-streaking and re-tested on extracellular enzyme detection plates. The plasmid DNA was isolated from Out- mutants restored to Out+ and used to transform E. coli DHI. This plasmid DNA was then prepared from E. coli DHI, analysed by restriction mapping and compared to the parent cosmid to investigate the nature of the deletion.

Cosmids with deletions which restored Out- mutants to Out+ were re-introduced into E. coli GJ342 by transformation and conjugally transferred into Out- mutants belonging to the group normally complemented by the parent cosmid. It was then possible to determine the effect of a particular deletion on all the mutants originally complemented by that parent cosmid.

5.8.1.2. Results and discussion

Cosmid cHIL159 (group c) was digested and re-ligated to promote re-circularisation without re-insertion of the released fragment. Table 5.4. shows

Legend for Figure 5.6.

GeneX is carried on a plasmid adjacent to the Tc resistance gene encoding Tet. Three hypothetical restriction enzymes (A, B and C) cut within the gene encoding Tet and within the DNA insert containing geneX.

1) Digestion with restriction enzyme A followed by religation might lead to the disruption of the Tet encoding gene and the loss of geneX.

2) Digestion with restriction enzyme B followed by religation might lead to the inactivation of the Tet encoding gene and disrupt geneX.

3) Digestion with restriction enzyme C followed by religation might lead to the inactivation of the Tet encoding gene. However, geneX would still be intact and carried on a smaller (deleted) DNA fragment.

Figure 5.6. Gene deletions

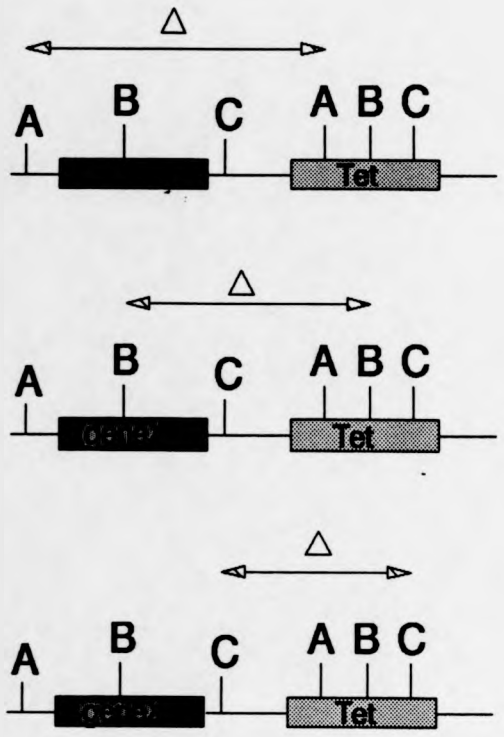


Table 5.4. Subcloning the C-group region by generating deletions in cosmid cHIL159 using various restriction enzymes

Restriction enzyme	Complementation of RJP159
EcoRI	-
EcoRV	-
SalI	-
ClaI	-
AvaI	-
SphI	-
HindIII	+
BglII	+

Legend

The symbols for complementation were described in Table 5.3.

the restriction endonucleases used and the effect upon complementation of deleted cosmids after re-introduction back into RJP159. Digestion with BglII (leading to pHIL159/2) and HindIII (leading to pHIL159/1) resulted in plasmids which could still complement RJP159 and which were also Tc^S. This indicated that a deletion event had taken place but not to an extent where the ability to complement this particular mutant had been impaired. The construction of pHIL159/1 and pHIL159/2 is illustrated in Figure 5.7. When cut with the other restriction enzymes indicated, cHIL159 must have been deleted leading to the loss or inactivation of the out+ genes required for the complementation of RJP159.

Plasmid pHIL159/1 was then introduced into other Out- mutants which had been cured of any other pBR322 based plasmids. The ability of this plasmid to complement Out- mutants RJP254, RJP220, RJP249, RJP253 and RJP159 is shown in Table 5.5. This result indicates that the necessary information required to complement RJP159, RJP220, RJP254 and RJP253 was still present on this plasmid. However complementation of RJP249 was not achieved by this construct. This result suggested that the out+ gene necessary for the complementation of the RJP249 mutation had been deleted or inactivated. The implication of this result was that this gene was situated towards the 3' end of cHIL159 beyond or dissected by the HindIII site, thus defining another complementation group.

5.6.2. Cloning of a 3 kb HindIII - SalI fragment and its ability to complement RJP159 and RJP253

5.6.2.1. Results and discussion

By studying the restriction map of cHIL159 and the effects of

Legend for Figure 5.7

Cosmid cHIL159 was digested with either HindIII or BglII. After digestion the reaction mix was extracted with phenol and ligated under conditions to promote re-circularisation with loss of insert. Plasmids pHIL159/1 and pHIL159/2 were constructed using this approach.

Figure 5.7. Generating deletions in cosmid cHIL159

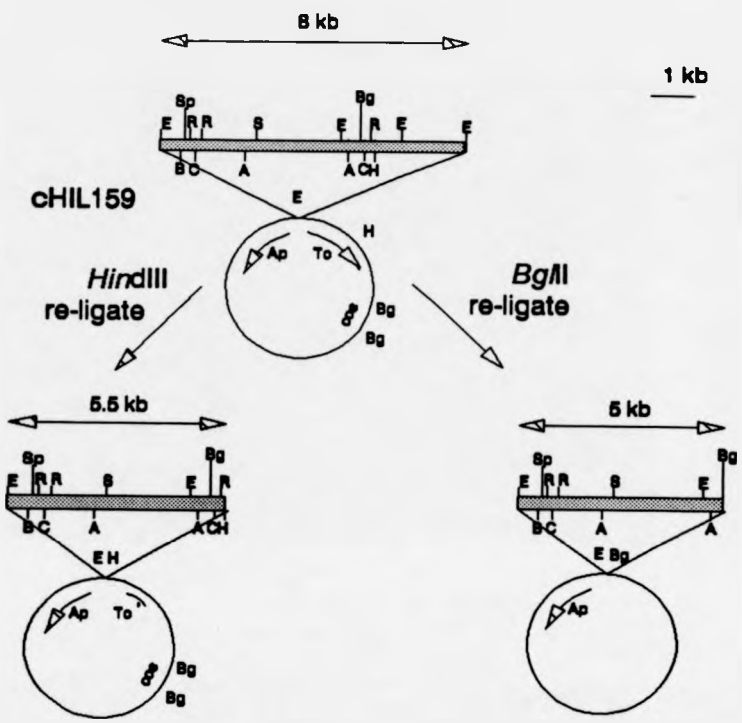


Table 5.5. Complementation pattern of pHIL159/1 when introduced into group B and group C Out- mutants

Mutant	Group	Complementation
RJP220	B	+
RJP159	C	+
RJP253	C	+
RJP249	C	-

Legend

The construction of pHIL159/1 is described in Figure 5.7. The complementation pattern was determined using plate assays and is discussed in the text.

deletions in this cosmid it was possible to pursue a more directed approach for subcloning out+ genes. For example, a deletion in cHIL159 using HindIII did not impair the ability of this construct to complement RJP159. However, deletions in this cosmid using other enzymes such as SalI or EcoRV did result in the loss of its ability to complement this mutant. It therefore followed that the gene necessary for complementation of RJP159 was situated between SalI/EcoRV and HindIII on this cosmid. This hypothesis was tested by isolating the HindIII to SalI fragment and investigating its ability to complement Out- mutants in this complementation region.

A random cloning approach was taken to subclone the HindIII to SalI fragment and is described in Figure 5.8. Plasmid pHIL159/1 was digested with HindIII and SalI and ligated with plasmid pBR322 which had also been digested with the same two restriction enzymes. Digested pBR322 had also been treated with calf intestinal phosphatase (section 2.22.). After ligation, the heterogeneous ligation mix was introduced into E. coli GJ342 by transformation (section 2.26.). Transformants were selected by their ability to grow on NA plates containing Ap Tc and Cm. Pooled transformants were then patch mated with RJP159. RJP159 plasmid recipients were selected on MMA containing Ap. RJP159 transconjugants were then screened on Pel and Cel detection plates for Out+ and on NA containing Tc to identify recombinant (Tc^S) pBR322 derivatives. Plasmid DNA from Out+ Tc^S RJP159 strains was prepared, and analysed to determine that the correct insert had been cloned. A plasmid (pHIL159/3) with the correct insert was used to transform the mobilising strain of E. coli GJ342. Plasmid pHIL159/3 was then mobilised into Ecc Out- mutants RJP254, RJP220, RJP249, RJP253 and RJP159. Transconjugants were screened for Out+ in order to identify complemented Out- mutants. The construction of plasmid (pHIL159/3) is illustrated in

Legend for Figure 5.8.

Plasmid pHIL159/1 was digested with HindIII and SalI in order to release the 3 kb HindIII - SalI fragment. This digested DNA was ligated with pBR322 which had been cut with the same two restriction enzymes and de-phosphatased. The isolation of pHIL159/3 is described in the text (section 5.8.2.).

Figure 5.8. Construction of pHIL159/3

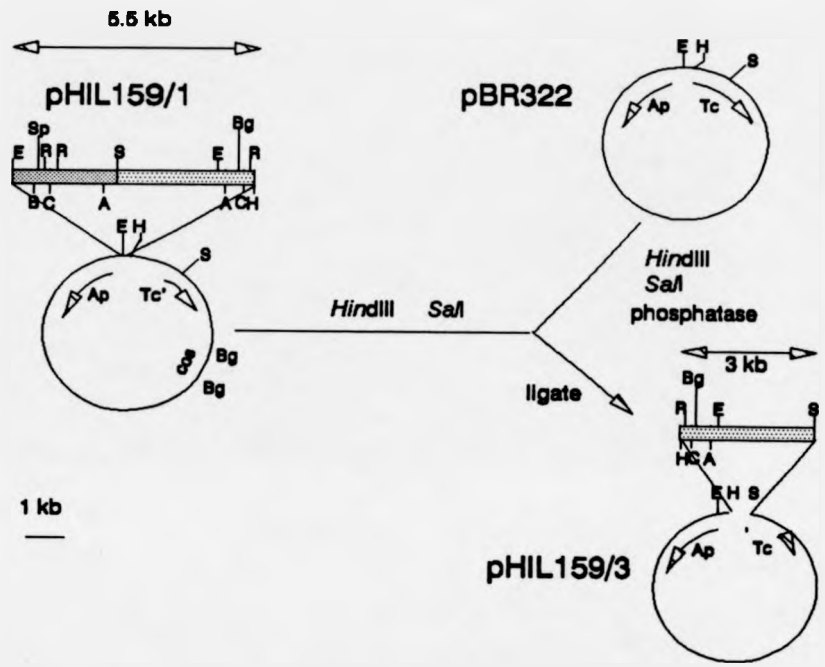


Figure 5.8. and its complementation pattern is shown in Table 5.6. Only RJP253 and RJP159 were complemented by this plasmid. This result indicated that the mutation of RJP249 was located between the 3' EcoRI site of cHIL220 and 5' of the SalI site of cHIL159, or, that the gene required for its complementation was disrupted by digestion with one these restriction enzymes.

5.7. A summary of the out gene cluster - restriction map and complementation groups

Three different cosmid groups which span a 12 kb fragment of DNA were isolated. These three overlapping cosmids restored the Out+ phenotype in specific groups of mutants. In total, 12 of the 14 EMS induced mutants were complemented by cosmids spanning this region. A large cosmid (cHIL208) which contained the 12 kb region and extra DNA did not fully complement any of the Out- mutants but partially complemented Out- mutants as determined by plate assays.

Figure 5.9. shows the six distinct groups as determined by complementation. Included in this restriction and complementation map is work that was carried out by D. Whitcombe (pers. comm.). This work includes the restriction mapping of the region complementing Out- mutant PR33 and also the subcloning of the out gene cluster upstream of PR33. Recently, the location of RJP249 at a position 3' of the HindIII site, at the extreme right hand side of the out gene cluster (as drawn), has been confirmed (N. Housby, pers. comm.).

On the basis of direct complementation, it was possible to identify at least six complementation groups within the 12 kb region. Recently, independent work in the laboratory of A. Chatterjee led to similar findings

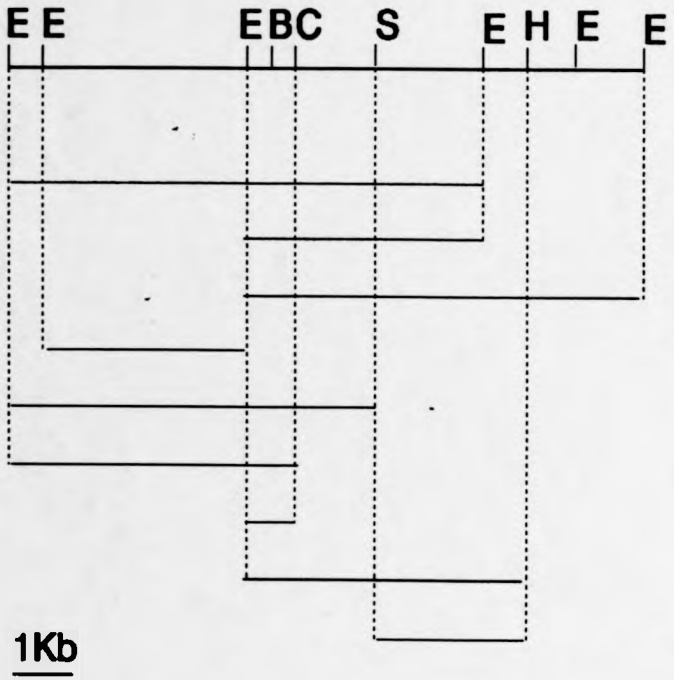
Table 5.6. Complementation pattern of pHIL159/3 when introduced into group B and group C Out- mutants

Mutant	Group	Complementation
RJP220	B	-
RJP159	C	+
RJP253	C	+
RJP249	C	-

Legend

The construction of pHIL159/3 is described in Figure 5.8. The complementation pattern was determined using plate assays and is discussed in the text.

Figure 5.9. Out complementation groups



Out mutant complemented					
RJP 190	RJP 122	PR 33	RJP 221	RJP 159	RJP 249
+	+	+	+	—	—
—	—	+	+	—	—
—	—	+	+	+	+
—	+	—	—	—	—
+	+	+	—	—	—
+	+	+	—	—	—
—	—	+	—	—	—
—	—	+	+	+	—
—	—	—	—	+	—

(Murata et al., 1990). Murata and co-workers isolated a large cosmid (pAK260) which complemented 31 Out- mutants of Ecc, Eca, and Ech. This cosmid also complemented all the Out- mutants of Ecc isolated in this laboratory (S. Wharam, pers. comm.). However, other workers studying Pel and Cel secretion in Ech suggest the existence of at least three non-clustered out loci (Ji et al., 1987 and 1989; Thurn and Chatterjee, 1985). This appears to contradict the findings of this work and recent work in the laboratory of A. Chatterjee. Even more recently, He et al. (1991a) isolated an out+ cosmid using Ech chromosomal DNA which restored transposon-induced Ech Out- mutants to Out+. A 12 kb fragment was isolated from this cosmid and four out complementation groups were identified. Significantly, the out+ genes contained on this 12 kb region enabled E. coli to efficiently secrete multiple pectolytic enzymes produced from cloned Ech genes. This is the first time that an Erwinia sp. Out secretion mechanism has been reconstructed in E. coli.

5.8. The ability of Out- mutants to rot potatoes when harbouring out+ cosmids

5.8.1. Introduction

In section 4.5. Out- mutants affected in their ability to rot potatoes under laboratory conditions were discussed. In general, Out- mutants displayed diminished levels of tuber maceration when inoculated into potato tubers. In this chapter the restoration of the Out+ phenotype has been demonstrated. I decided to test whether or not the restoration of the Out+ phenotype in Out- mutants (harbouring out+ cosmids) was concomitant with the return of wild-type potato maceration levels. Out- mutants harbouring

Table 5.7. Pathogenicity tests using mutants carrying out+ cosmids

Strain	Inoculum 1 (mm)	Inoculum 2 (mm)	Rot(average) (mm)	Degree of rot (% wt)
HC131	15	18	16.5	100
LE392	0	0	0	0
RJP200	16	14	15	91
RJP200(c)	17	18	17.5	106
RJP249	9	16	12.5	76
RJP249(c)	3	6	4.5	27
RJP253	8	0	4	24
RJP253(c)	3	0	1.5	9
RJP250	18	19	18.5	112
RJP250(c)	18	23	20.5	124
RJP221	0	6	3	18
RJP221(c)	3	4	3.5	21
RJP251	0	0	0	0
RJP254	0	4	2	12
RJP254(c)	8	9	8.5	52
RJP190	8	15	11.5	70
RJP190(c)	10	18	14	85
RJP159	8	3	5.5	33
RJP159(c)	7	0	3.5	21
RJP233	8	7	7.5	45
RJP233(c)	3	0	1.5	9
RJP208	11	3	7	42
RJP208(c)[a]	16	14	15	91
RJP122	15	3	9	55
RJP122(c)	5	18	11.5	70
RJP211	5	7	6	36
RJP220	6	3	4.5	27
RJP220(c)	20	15	17.5	106

Legend

(c) Mutant with cosmid which restores Out+ phenotype.
a Cosmid partially restored mutant to Out+.

This experiment was performed alongside that described in section 4.5 (Table 4.8). The pathogenicity tests for the Out- mutants and for the complemented (Out+) mutants are displayed together in this table so they can be directly compared.

out+ cosmids were inoculated into potato tubers (section 2.11.) in parallel with the Out- mutants alone (section 4.5.).

5.8.2. Results and discussion

The results from the soft rot experiment are given in Table 5.7. Rot diameters for single inoculums of each mutant (in duplicate) are shown by columns, inoculum 1 and inoculum 2. The average rot for each mutant and for some complemented mutants is shown. Complemented mutants are indicated by (c) where (c) represents a cosmid which fully restored the Out+ phenotype in that mutant. The final column compares the average rot of each mutant, complemented mutant and E. coli DH1, with the average rot achieved by Ecc HC131. The results indicate that one complemented Out- mutant, RJP220(cHIL220), had a comparable maceration level to the parent strain HC131. Some of the other mutants showed a smaller increase in maceration levels when carrying cosmids encoding out+ genes (e.g. RJP200(cHIL200), RJP254(cHIL159) and RJP208(cHIL208)) compared to the same mutants without the corresponding out+ cosmid. However, some mutants exhibited a decreased maceration level when harbouring the appropriate out+ cosmid (e.g. RJP249(cHIL159) and RJP159(cHIL159)). These results show that in some cases the level of potato rot increased for Out- mutants carrying out+ cosmids, suggesting that the Out+ phenotype might be an important pathological trait in this bacterium. However, these results can only be considered as preliminary findings and some of these results are difficult to explain. The experiment needs to be refined in order to tightly control factors such as inoculum size, the method of determining maceration levels and tuber variability. In this particular experiment it might also have been interesting to monitor the maintenance of the out+ cosmids in the Out-

mutants when inoculated into potato tubers. It might be expected that a cosmid restoring an Out+ phenotype would incur a selective advantage on an Out- mutant when growing in a potato tuber.

In a separate piece of work, Murata et al. (1990) isolated a cosmid which complemented Ecc, Eca and Ech Out- mutants including the Ecc Out- mutants isolated in this laboratory (S. Wharam pers. comm). This cosmid (pAKC260) restored the ability of two Ecc Out- mutants (AC5021 and AC5088) to rot potatoes under laboratory conditions (Murata et al., 1990).

5.9. The restoration of bacteriophage sensitivity in RJP190 when harbouring a out+ cosmid

5.9.1. Introduction

The phage resistant phenotype of Out- mutant RJP190 has been described (section 4.4.). Ecc bacteriophages (ϕKP, ϕD-2, ϕ565 and ϕ575) were spotted onto bacterial lawns of RJP190 and RJP190(cHIL190) in order to investigate the effect of the out+ cosmid on the phage resistant phenotype of RJP190. The ability of RJP190 (cHIL190) to adsorb bacteriophage ϕD-2 was also investigated. These experiments were performed in parallel with those described in section 4.4.

5.9.2. Results and discussion

The results of a bacteriophage efficiency of plating test are shown in Figure 5.10. All four bacteriophages infected HC131. However, ϕD-2 and ϕ565 appeared to be of a lower titre than ϕ575 and ϕKP. On RJP190, both ϕKP and ϕ575 infected to the same level as on HC131. However, ϕD-2 and ϕ565 did not infect even at the 10^{-2} dilution on Out- mutant RJP190. This effect

Figure 5.10. Cosmid cHIL190 restores the sensitivity of RJP190 to bacteriophages ∅565 and ∅D-2

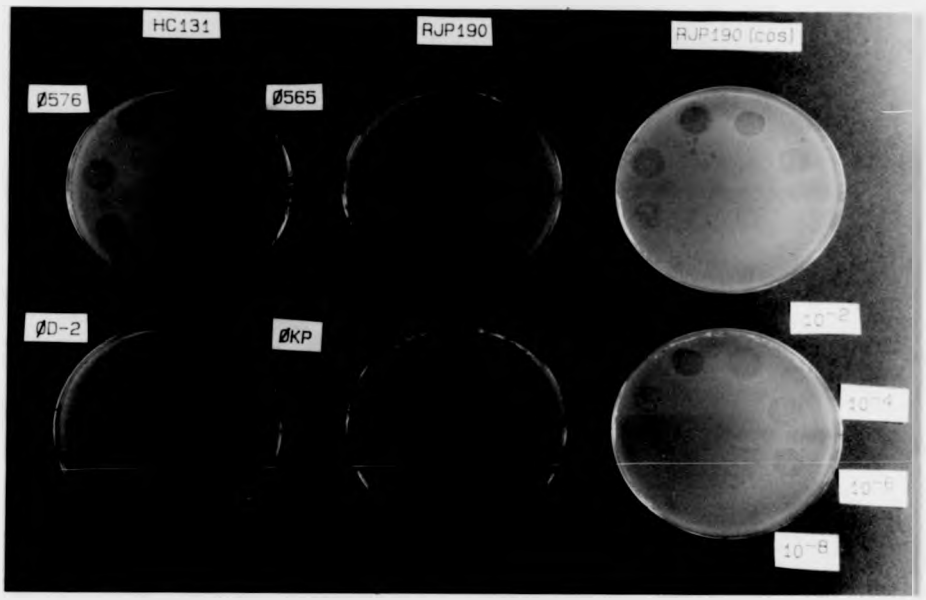


Figure legend

Serial dilutions of bacteriophage stocks were spotted onto lawns of the above strains as illustrated in the photograph. The results from this experiment are discussed in the text (5.9.2.)

can clearly be seen by the absence of a clear spot at the 10^{-2} position on the bacterial lawn of RJP190. When infecting RJP190(cHIL190), both ϕD-2 and ϕ565 infected to the same levels as on HC131. This result strongly suggests that as well as restoring the Out+ phenotype in RJP190, the out+ cosmid (cHIL190) also restores the sensitivity of RJP190 to Ecc bacteriophages ϕD-2 and ϕ565.

The results of the bacteriophage ϕD-2 adsorption assay are shown in Table 5.8. The results show that RJP190 failed to adsorb ϕD-2 whereas HC131 removed 83% pfu from the suspension. Another Out- mutant, RJP250, was also efficient at adsorbing ϕD-2. Out mutant RJP190 carrying cHIL190 (which restores Out+ and ϕD-2 and ϕ565 sensitivity) adsorbed the same number of pfu as HC131. The E. coli DH1 control did not appear to adsorb the Ecc bacteriophage ϕD-2.

The results from the above experiments suggest that an Out protein (which RJP190 lacks) might be simultaneously involved in the sensitivity of Ecc to bacteriophages ϕD-2 and ϕ565, presumably by being involved in the adsorption of these two bacteriophages. It is tempting to speculate that ϕD-2 and ϕ565 are entering this bacterium through a route that involves adsorption to the Ecc Out secretion mechanism.

Table 5.8. Adsorption assays using bacteriophage ØD-2

Strain	ØD-2 t=0(min) pfu/ml	ØD-2 t=25(min) pfu/ml	ØD-2 pfu/ml (adsorbed)	Adsorption (%)
HC131	6.0×10^{10}	1.0×10^{10}	5.0×10^{10}	83
DH1	6.0×10^{10}	7.0×10^{10}	0	0
RJP190	6.0×10^{10}	7.0×10^{10}	0	0
RJP190 (cHIL190)	6.0×10^{10}	1.0×10^{10}	5.0×10^{10}	83
RJP253	6.0×10^{10}	2.0×10^{10}	4.0×10^{10}	67

Legend

This experiment was performed in parallel with the adsorption experiment described in section 4.4.3., the results of which are summarised in Table 4.7. In this experiment Out- mutant RJP190 carrying the out+ cosmid was also tested for its ability to adsorb ØD-2. The final column in this table (% adsorption) shows how efficient a particular strain was at removing (adsorbing) ØD-2 from the bacterial supernatant under the conditions used. These results are discussed in the relevant text (section 5.9.2.).

CHAPTER 6

ANALYSIS OF *out* GENES AND THEIR ENCODED PROTEINS

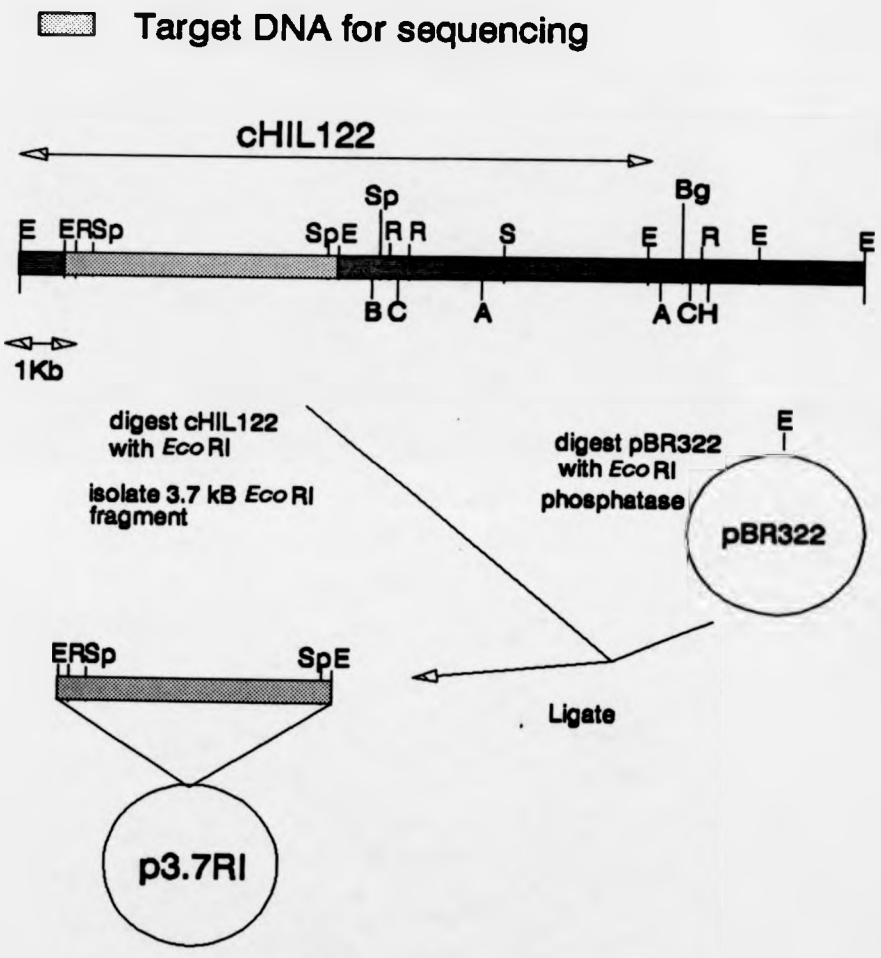
6.1. Introduction

In the previous chapter it was demonstrated that a 12 kb EcoRI fragment encoded functions which restored the Out+ phenotype in 12 of the 14 EMS Out- mutants. Furthermore, it was possible to assign particular mutants to specific regions of this cluster of out+ genes. In order to gain an insight into the nature of these genes at a molecular level the DNA sequence of a region containing out genes was determined. In this chapter the DNA sequence data from the 3.7 kb EcoRI fragment, as illustrated in Figure 6.1., will be presented. This region was found to overlap with sequence data generated independently by other workers in this laboratory (D. Whitcombe and M. Gibson). The analysis of this contiguous DNA sequence and the predicted protein products will be presented.

6.2. Generation of clones for sequencing

The method used for the generation of clones for sequencing is described in sections 2.27.1. to 2.27.11. (from Bankier et al., 1986). Plasmid pBR322 containing the 3.7 kb EcoRI fragment (p3.7RI) was prepared from E. coli DH1 using the 'midi-prep' procedure (section 2.18.3.). The 3.7 kb EcoRI fragment was released from the pBR322 vector by digestion to completion with EcoRI. This digestion would have resulted in two DNA fragments of similar molecular weight (3.7 kb and 4.4 kb). In order to simplify the task of separating these two bands on an agarose gel, the restriction enzyme PvuII was also added to the restriction digest reaction. This restriction enzyme cuts within the vector DNA only, thus generating two fragments of 2.3 kb and 2.1 kb. The 3.7 kb DNA fragment was collected using the 'Biorad' electroelutor apparatus (section 2.27.2.). The 3.7 kb fragment was circularised by self-ligation (section 2.27.3.) and sonicated

Figure 6.1. Sequencing the 3.7 kb *Eco*RI fragment



(section 2.27.4.). The sonicated DNA was then end-repaired (section 2.27.5.) and size fractionated (section 2.27.6.). The size fractionated/end-repaired DNA was ligated into M13 mp8 (section 2.27.8.) as described in (section 2.27.9.). The ligation mix was introduced into E. coli TG1 by transformation (section 2.27.10.). Plugs of top agar containing 'white' (recombinant) plaques were removed using a Pasteur pipette and stored at 4°C in 200 ul of TMG (phage) buffer.

6.2.1. Results

A series of control ligations were set up in order to identify the source of any problems and are shown in Table 6.1. This protocol resulted in the production of 204 'white' (recombinant) M13 plaques which were used to prepare DNA template for sequencing.

6.3. Sequencing and compilation of clones

6.3.1. Introduction

Template DNA for sequencing was made from the white 'recombinant' plaques as described in section 2.27.11. These templates were sequenced using 'Sequenase' as described in section 2.27.12. DNA sequences from gels were read into an IBM personal computer using a 'sonic digitiser'. Sequences were manipulated and analysed using the Beckman 'Microgenie' programmes (Queen and Korn, 1984). The 'Data bank' programme was used to eliminate vector M13 mp8 and pBR322 contaminating DNA sequences. Sequences were merged using the 'Shotgun merge' programme to produce a contiguous sequence. The fragment was sequenced in both strands at least once and in some regions up to 12 times. Compressions were eliminated by substituting dGTP with dITP

Table 6.1. Ligation into M13 mp8

	Tube number					
	1	2	3	4	5	6
Vector/buffer[1]	8µl	8µl	8µl	8µl	8µl	8µl
DNA	1µl	2µl	3µl	-	-	-
Alu1 cut lambda (10ng/µl)	-	-	-	1µl	-	-
T4 DNA ligase[2]	1µl	1µl	1µl	1µl	1µl	-
	Resulting plaques per plate					
blue plaques[3]	10	14	9	14	6	-
white plaques[3]	33	51	120	>1000	1	-

1 Vector buffer mix was made by mixing:

6µl (10x) ligation buffer

6µl (120ng) *Sma1* cut M13

41µl water

2 1µl = 100 units of activity

3 Number of plaques per plate after transformation into *E. coli* TG1

in the sequencing reactions as described by the manufacturers of 'Sequenase'. The complementary strand of single stranded regions was obtained by using extended gel runs of up to 6 hr in total.

For one region that had no clones present for one strand, plasmid sequencing was carried out (section 2.27.13.). A primer was designed to anneal to the complementary strand allowing sequencing of this single stranded region. The plasmid used was pBR322 containing the 3.7 kb region. The DNA sequence of the primer was 3'-ACGTATTGGGGCAACGC-5' which annealed at position 3455-3474.

6.3.2. Results

The 3.7 kb EcoRI fragment was found to contain 3822 bp when its nucleotide sequence was determined. The nucleotide sequence of a region of the *out* gene cluster, including the 3.8 kb EcoRI fragment, is shown in Figure 6.2. The sequence includes residues 1-965 at the 5' end (sequenced by M. Gibson) and residues 4792-5673 at the 3' end (sequenced by D. Whitcombe). Residues 966-4791 between the EcoRI were sequenced as part of this work. DNA spanning these EcoR1 sites was also sequenced.

6.4. Analysis of the DNA sequence

6.4.1. Introduction

The DNA sequence was subjected to the 'find possible coding regions' programme which is contained in the 'Analysis' menu of Microgenie (Queen and Korn, 1984), in order to identify potential coding regions. The Staden protein analysis package (using an IBM computer) was used to identify potential transcription start sites (Staden, 1978).

Figure 6.2. Nucleotide sequence of *out* genes

```
        10         20         30         40         50         60
TTCGAGGATAACGATCTCGCCTGTCATTAAATGGCATGGATTTACGTGATCGGGATCAGG

        70         80         90        100        110        120
CACAGCAGGCGATGGCGCAACTGGCAGGGATGAGCAAATTTAATTTGACCGTCGAGCGTG

       130        140        150        160        170        180
ATGGTCAACAGCAGGATATATATCTGGCACTGGATGGAGACCACTAATTTGTTTAGCAAG

       190        200        210        220        230        240
GGACAGGGATTTTTTAAACGTCAGGTTTTTTCGAAGAATAAAAACCAATGGCTTGGTCAG

       250        260        270        280        290        300
GTACGCCGCAAGAGCATGCTATTGCTCAGCGGGAGTGTTCTGCTGATGGCGTCATCATTG
  M  R  R  K  S  M  L  L  L  S  G  S  V  L  L  M  A  S  S  L
OutD »        EcoRI
       310        320        330        340        350        360
GCGTGGAGCGCTGAATTCTCCGCCAGTTTTAAAGGCACCGATATTCAGGAGTTTATTAAT
  A  W  S  A  E  F  S  A  S  F  K  G  T  D  I  Q  E  F  I  N
                ↑
       370        380        390        400        410        420
ACCGTCAGTAAGAATCTGAATAAGACGGTCATTATCGATCCGTCGGTCAGCGGAACGATT
  T  V  S  K  N  L  N  K  T  V  I  I  D  P  S  V  S  G  T  I

       430        440        450        460        470        480
ACCGTACGCAGCTATGACATGATGAATGAAGAACAGTATTACCAGTTCTTCCTGAGCGTG
  T  V  R  S  Y  D  M  M  N  E  E  Q  Y  Y  Q  F  F  L  S  V

       490        500        510        520        530        540
CTGGATGTCTATGGGTTCACCGTGATTCCGATGGATAACAACGTTCTGAAAATCATTCGC
  L  D  V  Y  G  F  T  V  I  P  M  D  N  N  V  L  K  I  I  R

       550        560        570        580        590        600
TCGAAGGATGCGAAATCGACGTCTATGCCGCTGGCCACTGATCGACAGCCGGGCATTGGT
  S  K  D  A  K  S  T  S  M  P  L  A  T  D  R  Q  P  G  I  G

       610        620        630        640        650        660
GATGAGGTTGTGACGCGCGTTGTGCCCGTCAACAACGTCGCCGCGCGTGATTTCGGACGC
  D  E  V  V  T  R  V  V  P  V  N  N  V  A  A  R  D  F  G  R

       670        680        690        700        710        720
TCGTCGCGAGTTGAACGACAACGCGTGGCGTGGGACGTGTGGCGATTACGAACCTGCGAA
  S  S  R  V  E  R  Q  R  V  A  W  D  V  W  R  L  R  T  C  E

       730        740        750        760        770        780
CGTCGTCGTGATGACTGGCCGCGCGGGGTGATCCACGCGGTGATGACGATTGTCGAACGG
  R  R  R  D  D  W  P  R  G  V  I  H  A  V  M  T  I  V  E  R

       790        800        810        820        830        840
GTTGATCAGACGGGCGATCGGAATGTGACCACGATACCGCTGTCTTACGCTTCCTCAACG
  V  D  Q  T  G  D  R  N  V  T  T  I  P  L  S  Y  A  S  S  T

       850        860        870        880        890        900
GAAGTGGTGAAGATGGTGAATGAGCTGAACAAGATGGATGAGAAATCCGCCTTGCCCGGC
  E  V  V  K  M  V  N  E  L  N  K  M  D  E  K  S  A  L  P  G
```

```
      910        920        930        940        950        960
ATGTTGACCGCCAACGTAGTGGCTGACGAGCGAACCAACTCGGCTGCTGGCTTCGGCGAG
 M  L  T  A  N  V  V  A  D  E  R  T  N  S  A  A  G  F  G  E
   EcoR1
      970        980        990       1000       1010       1020
CCGAATTCCCGCCAGCGTGTGATTGATATGGTCAAGCAGCTCGATCGCCAGCAGGCGGTA
 P  N  S  R  Q  R  V  I  D  M  V  K  Q  L  D  R  Q  Q  A  V

     1030       1040       1050       1060       1070       1080
CAGGGCAACACCAAAGTTATCTACCTCAAATACGCCAAAGCCGCCGATCTGGTCGAAGTG
 Q  G  N  T  K  V  I  Y  L  K  Y  A  K  A  A  D  L  V  E  V

     1090       1100       1110       1120       1130       1140
CTCACCGGTGTCGGCGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
 L  T  G  V  G  D  S  I  Q  T  D  Q  Q  N  A  L  P  A  L  R

     1150       1160       1170       1180       1190       1200
AAAGACATTTCGATTAAGGCACACGAACAAACCAACTCGCTGATTGTGAATGCCGCACCG
 K  D  I  S  I  K  A  H  E  Q  T  N  S  L  I  V  N  A  A  P
                                      EcoRV
     1210       1220       1230       1240       1250       1260
GACATCATGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAGGTG
 D  I  M  R  D  L  E  Q  V  I  A  Q  L  D  I  R  R  P  Q  V

     1270       1280       1290       1300       1310       1320
TTGGTCGAAGCGATCATCGCGGAAGTACAGGATGCCGATGGCATGAATCTGGGCGTTCAA
 L  V  E  A  I  I  A  E  V  Q  D  A  D  G  M  N  L  G  V  Q

     1330       1340       1350       1360       1370       1380
TGGGCGAATAAAAATGCTGGTGTAACGCAATTCACGAATACGGGATTACCGATCACAACA
 W  A  N  K  N  A  G  V  T  Q  F  T  N  T  G  L  P  I  T  T

     1390       1400       1410       1420       1430       1440
ATGATGGCAGGGGCAGACCAGTTTCGGCGTGATGGAACACTTGGTACGGCCGCGACAACG
 M  M  A  G  A  D  Q  F  R  R  D  G  T  L  G  T  A  A  T  T
                                                  SphI
     1450       1460       1470       1480       1490       1500
GCACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCTG
 A  L  G  G  F  N  G  I  A  A  G  F  Y  Q  G  N  W  G  M  L

     1510       1520       1530       1540       1550       1560
ATGACGGCACTGTCCAGCAACAGTAAAAACGATATTCTGGCAACGCCCAGTATTGTGACG
 M  T  A  L  S  S  N  S  K  N  D  I  L  A  T  P  S  I  V  T

     1570       1580       1590       1600       1610       1620
CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG
 L  D  N  M  E  A  T  F  N  V  G  Q  E  V  P  V  L  A  G  S

     1630       1640       1650       1660       1670       1680
CAGACGACATCCGGTGATAACGTTTTCCAAACCGTGGAACGTAAAACGGTCGGTATCAAG
 Q  T  T  S  G  D  N  V  F  Q  T  V  E  R  K  T  V  G  I  K

     1690       1700       1710       1720       1730       1740
CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
 L  K  V  K  P  Q  I  N  E  G  D  S  V  L  L  E  I  E  Q  E
```

```
      1750       1760       1770       1780       1790       1800
GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
 V  S  S  V  A  D  A  A  S  S  S  S  T  N  L  G  A  T  F  N

      1810       1820       1830       1840       1850       1860
ACGCGTACCGTGAATAACGCGGTACTGGTCAGCAGCGGCGACACCGTAGTGGTGGGCGGT
 T  R  T  V  N  N  A  V  L  V  S  S  G  D  T  V  V  V  G  G

      1870       1880       1890       1900       1910       1920
TTGTTGGATAAAAGTACCAATGAGTCTGCAAATAAAGTGCCCCTTTTGGGCGATATTCCC
 L  L  D  K  S  T  N  E  S  A  N  K  V  P  L  L  G  D  I  P

      1930       1940       1950       1960       1970       1980
GTGCTGGGATATTTGTTCCGTTCCAACAGCACGGAAACGAAAAAGCGTAACCTGATGCTG
 V  L  G  Y  L  F  R  S  N  S  T  E  T  K  K  R  N  L  M  L

      1990       2000       2010       2020       2030       2040
TTTATCCGTCCTTCCATTATTCGCGATCGCAGCCAATTCCAGAGCGCCTCTGCCAGTAAG
 F  I  R  P  S  I  I  R  D  R  S  Q  F  Q  S  A  S  A  S  K

      2050       2060       2070       2080       2090       2100
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGGGAGGG
 Y  N  S  F  S  A  E  E  N  K  Q  R  N  V  S  N  G  E  G  G

      2110       2120       2130       2140       2150       2160
CTTCTGGATAACGATTTGCTGCGCTTGCCGGAAGGTGGAAATGCCTATACGTTCCGTCAG
 L  L  D  N  D  L  L  R  L  P  E  G  G  N  A  Y  T  F  R  Q

      2170       2180       2190       2200       2210       2220
GTTCAGTCCTCCATTGTGGCGTTTTATCCGGCGGGCGGGAAATGAGTGACGTTGCCTCCC
 V  Q  S  S  I  V  A  F  Y  P  A  G  G  K  *
                                   OutE »  M  S  D  V  A  S

     2230       2240       2250       2260       2270       2280
AGATTATAGAGTTACGCCCCATACTGCCTTTTGCCTATGCACGATCGCAGCAAATTCTGC
 Q  I  I  E  L  R  P  I  L  P  F  A  Y  A  R  S  Q  Q  I  L

      2290       2300       2310       2320       2330       2340
TGTTGCAGAGGGAAAATGACGCGAGCTTACAGACGATTTGCGTCGCGCAAACGCCGCCAG
 L  L  Q  R  E  N  D  A  S  L  Q  T  I  C  V  A  Q  T  P  P

      2350       2360       2370       2380       2390       2400
CCGCTTTGCTGGAAGCGCGTCGGATTGCAGGCTGTTCGCTCAGGATTGAGCGCGTTACGG
 A  A  L  L  E  A  R  R  I  A  G  C  S  L  R  I  E  R  V  T

      2410       2420       2430       2440       2450       2460
ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGACTCGGAAGAAGCGCGCC
 D  E  E  F  E  R  Q  L  V  I  S  Y  Q  R  D  S  E  E  A  R

      2470       2480       2490       2500       2510       2520
GTATGATGGAGGACATTGGTAATGAGATGGACTTCTATACGCTGGTGGAAGAACTACCAG
 R  M  M  E  D  I  G  N  E  M  D  F  Y  T  L  V  E  E  L  P

      2530       2540       2550       2560       2570       2580
ATAGCGATGACTTGCTCGATGCCGATGACGACGCGCCGATTATCCGCCTCATCAACGCCA
 D  S  D  D  L  L  D  A  D  D  D  A  P  I  I  R  L  I  N  A
```

```
      2590       2600       2610       2620       2630       2640
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATATTCATATCGAAACGTATGAGCGCT
M  L  T  E  A  I  K  N  K  A  S  D  I  H  I  E  T  Y  E  R

      2650       2660       2670       2680       2690       2700
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
Y  L  L  I  R  F  R  V  D  G  V  L  R  E  I  L  R  P  Q  R

      2710       2720       2730       2740       2750       2760
AGCTGGCTTCGCTGCTGGTGTCGCGTATCAAGGTCATGGCAAAGCTGGATATTGCGGAAA
K  L  A  S  L  L  V  S  R  I  K  V  M  A  K  L  D  I  A  E

      2770       2780       2790       2800       2810       2820
AGCGTGTCCCGCAGGATGGACGTATGGCGCTGCGAGTGGGGGGGCGGGCGATTGATGTCC
K  R  V  P  Q  D  G  R  M  A  L  R  V  G  G  R  A  I  D  V

      2830       2840       2850       2860       2870       2880
GTGTCTCCACGCTGCCGTCGAACTACGGCGAGCGCGTCGTGCTGCGTTTGCTGGATAAAA
R  V  S  T  L  P  S  N  Y  G  E  R  V  V  L  R  L  L  D  K

      2890       2900       2910       2920       2930       2940
ACAGCGTTAAGCTCGATCTTGAGCTGCTGGGGATGTCGGAACGCAATCGACAACTGCTCG
M  S  V  K  L  D  L  E  L  L  G  M  S  E  R  N  R  Q  L  L

      2950       2960       2970       2980       2990       3000
ACAGCCTGATTCATCGTCCTCATGGCATTATCCTGGTCACCGGCCCGACAGGCTCGGGGA
D  S  L  I  H  R  P  H  G  I  I  L  V  T  G  P  T  G  S  G

      3010       3020       3030       3040       3050       3060
AAAGTACCACGCTTTACGCCGCGCTCAGCCGCCTGAATGCTTCGGAACGTAACATCATGA
K  S  T  T  L  Y  A  A  L  S  R  L  N  A  S  E  R  N  I  M

      3070       3080       3090       3100       3110       3120
CGGTGGAAGATCCCATCGAGTATGAACTGGAGGGTATCGGGCAAACGCAGGTCAACACCA
T  V  E  D  P  I  E  Y  E  L  E  G  I  G  Q  T  Q  V  N  T

      3130       3140       3150       3160       3170       3180
AGGTCGATATGACGTTTGCCCGCGGGCTGCGTGCCATTCTGCGTCAGGACCCGGACGTCG
K  V  D  M  T  F  A  R  G  L  R  A  I  L  R  Q  D  P  D  V

      3190       3200       3210       3220       3230       3240
TGCTGGTGGGGGAAATTCGTGATGGTGAAACGGCGCAGATTGCCGTGCAGGCCTCGTTGA
V  L  V  G  E  I  R  D  G  E  T  A  Q  I  A  V  Q  A  S  L

      3250       3260       3270       3280       3290       3300
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGCGCGCTGTCCCGTT
T  G  N  L  V  L  S  T  L  H  T  N  S  A  L  G  A  L  S  R

      3310       3320       3330       3340       3350       3360
TGCAGGATATGGGCGTTGAGCCTTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCGC
L  Q  D  M  G  V  E  P  F  L  L  S  T  S  L  L  G  V  L  A

      3370       3380       3390       3400       3410       3420
AGCGTCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
Q  R  L  V  R  T  L  C  S  D  C  S  Q  P  Q  P  V  D  P  V
```

```
      3430       3440       3450       3460       3470       3480
AGGCTGAACAGATGGGGATCGCGCCCGGTACGCTACTGCATAACCCCGTTGGCTGTCCGC
Q  A  E  Q  M  G  I  A  P  G  T  L  L  H  N  P  V  G  C  P

      3490       3500       3510       3520       3530       3540
AGTGTAGCTTTACCGGCTACCGGGGACGTATCGGCATTCATGAACTGGTGCTGATTAATG
Q  C  S  F  T  G  Y  R  G  R  I  G  I  H  E  L  V  L  I  N

      3550       3560       3570       3580       3590       3600
ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
D  D  V  R  A  A  I  H  R  S  D  G  E  M  A  I  A  Q  I  L

      3610       3620       3630       3640       3650       3660
GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
G  G  S  R  T  T  I  R  Q  D  G  L  N  K  V  L  A  G  L  T

      3670       3680       3690       3700       3710       3720
CCTGGGAAGAAGTGATCCGCGTAACCAAAGAGGAATGATATGGCACAGTACCACTATCAG
                                                M  A  Q  Y  H  Y  Q
T  W  E  E  V  I  R  V  T  K  E  E  *  OutF »

      3730       3740       3750       3760       3770       3780
GCGCTGGATGCGCAGGGGAAAAAATGCCGTGGCACTCAGGAGGCCGACTCTGCCAGACAG
 A  L  D  A  Q  G  K  K  C  R  G  T  Q  E  A  D  S  A  R  Q

      3790       3800       3810       3820       3830       3840
GCGCGTCAGCTATTGCGGGAGCGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC
 A  R  Q  L  L  R  E  R  G  L  V  P  L  S  V  D  E  N  R  G

      3850       3860       3870       3880       3890       3900
GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC
 D  Q  Q  K  S  G  S  T  G  L  S  L  R  R  K  I  R  L  S  T

      3910       3920       3930       3940       3950       3960
TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTGGCGGCGTCGATGCCGCTG
 S  D  L  A  L  L  T  R  Q  L  A  T  L  V  A  A  S  M  P  L

      3970       3980       3990       4000       4010       4020
GAAGAAGCGCTGGATGCGGTGGCAAAACAGAGCGAAAAGCCACATCTGAGCCAGTTGATG
 E  E  A  L  D  A  V  A  K  Q  S  E  K  P  H  L  S  Q  L  M

      4030       4040       4050       4060       4070       4080
GCGGCGGTGCGCAGCAAGGTCATGGAAGGTCACTCGCTGGCCGATGCCATGAAATGCTTT
 A  A  V  R  S  K  V  M  E  G  H  S  L  A  D  A  M  K  C  F

      4090       4100       4110       4120       4130       4140
CCCGGCAGCTTTGAACGGCTGTACTGCGCGATGGTTGCCGCAGGCGAGACATCCGGTCAC
 P  G  S  F  E  R  L  Y  C  A  M  V  A  A  G  E  T  S  G  H

      4150       4160       4170       4180       4190       4200
CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGCCAGCAGATGCGCAGC
 L  D  A  V  L  N  R  L  A  D  Y  T  E  Q  R  Q  Q  M  R  S

      4210       4220       4230       4240       4250       4260
CGCATCCAGCAGGCGATGATTTACCCGTGCGTGTTAACCGTGGTCGCGATTGCCGTGGTC
 R  I  Q  Q  A  M  I  Y  P  C  V  L  T  V  V  A  I  A  V  V
```

```
      4270       4280       4290       4300       4310       4320
AGCATTTTGCTGTCCGTCGTGGTGCCGAAAGTTGTTGAGCAATTTATTCATATGAAGCAG
 S  I  L  L  S  V  V  V  P  K  V  V  E  Q  F  I  H  N  K  Q

      4330       4340       4350       4360       4370       4380
GCGCTGCCGCTTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
 A  L  P  L  S  T  R  V  L  M  G  M  S  D  A  V  R  T  F  G

      4390       4400       4410       4420       4430       4440
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCCGCGTGATGCTGCGGCAG
 P  W  M  L  L  A  L  L  A  G  F  M  A  F  R  V  M  L  R  Q

      4450       4460       4470       4480       4490       4500
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA
 E  K  R  R  V  S  F  H  R  R  L  L  H  L  P  L  I  G  R  I

      4510       4520       4530       4540       4550       4560
GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG
 A  R  G  L  N  T  A  R  Y  A  R  T  L  S  I  L  N  A  S  A

      4570       4580       4590       4600       4610       4620
GTGCCGCTGTTACAGGCGATGCGCATTAGCGGTGATGTCATGAGCAATGACTATGCCCGC
 V  P  L  L  Q  A  M  R  I  S  G  D  V  M  S  N  D  Y  A  R

      4630       4640       4650       4660       4670       4680
CATCGGCTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCAGCCTGCACAAAGCGCTG
 H  R  L  S  L  A  T  D  A  V  R  E  G  V  S  L  H  K  A  L

      4690       4700       4710       4720       4730       4740
GAACAGACGGCGCTCTTCCCCCCAATGATGCGCCATATGATCGCCAGCGGTGAACGCAGC
 E  Q  T  A  L  F  P  P  M  M  R  H  M  I  A  S  G  E  R  S
            SphI                                      EcoRI
      4750       4760       4770       4780       4790       4800
GGTGAGCTGGACAGCATGCTGGAGCGGGCGGCGGATAATCAGGATCGGGAATTCAGTTCG
 G  E  L  D  S  M  L  E  R  A  A  D  N  Q  D  R  E  F  S  S

      4810       4820       4830       4840       4850       4860
CAGATGACGCTGGCGCTGGGGCTATTTGAGCCTCTGCTGGTGGTTAGCATGGCGGCGGTG
 Q  M  T  L  A  L  G  L  F  E  P  L  L  V  V  S  M  A  A  V

      4870       4880       4890       4900       4910       4920
GTGTTGTTTATCGTACTCGCGATTTTACAACCGATTCTGCAACTGAATACGTTAATGAGT
 V  L  F  I  V  L  A  I  L  Q  P  I  L  Q  L  N  T  L  M  S

      4930       4940       4950       4960       4970       4980
TCGTAAGAGAACCAACAATTGAATGAGGAAAAGTAAGGATGCAACAGTCTCAGCGTGGTT
 S  *
                                  OutG »  M  Q  Q  S  Q  R  G

      4990       5000       5010       5020       5030       5040
GTGGACAAAATAGTTATGGTCAGAGCGGTTATCGTCAGCGTGGTTTTACCCTGTTGGAGA
 C  G  Q  N  S  Y  G  Q  S  G  Y  R  Q  R  G  F  T  L  L  E

      5050       5060       5070       5080       5090       5100
TTATGGTGGTCATCGTCATTCTCGGCGTACTGGCGAGTCTGGTGGTGCCCAATCTGATGG
 I  M  V  V  I  V  I  L  G  V  L  A  S  L  V  V  P  N  L  M
```

```
      5110       5120       5130       5140       5150       5160
GGAATAAGGAAAAGGCGGATCGGCAAAAAGCCGTCAGCGATATTGTTTCTCTGGAAAGCG
G  N  K  E  K  A  D  R  Q  K  A  V  S  D  I  V  S  L  E  S

      5170       5180       5190       5200       5210       5220
CACTCGACATGTACAAGCTGGATAACAACCGCTATCCGTCCACAGAGCAGGGACTGAAGG
A  L  D  M  Y  K  L  D  N  N  R  Y  P  S  T  E  Q  G  L  K

      5230       5240       5250       5260       5270       5280
CGCTGGTAACGAAGCCGACGGTACAGCCGGAACCACGTAACTACCCAGCCGATGGCTATA
A  L  V  T  K  P  T  V  Q  P  E  P  R  N  Y  P  A  D  G  Y
                 BamHI
      5290       5300       5310       5320       5330       5340
TTCGCCGCTTGCCGCAGGATCCGTGGGGTACGGACTATCAACTGCTGAACCCTGGCCAGC
I  R  R  L  P  Q  D  P  W  G  T  D  Y  Q  L  L  N  P  G  Q
            EcoRV                          SphI
      5350       5360       5370       5380       5390       5400
ACGGCAAACTGGATATCTTCTCTCTGGGGCCAGATGGCATGCCGGGAACGAAGATGACAT
H  G  K  L  D  I  F  S  L  G  P  D  G  M  P  G  T  K  M  T

      5410       5420       5430       5440       5450       5460
CGGCAACTGGAATCTTGATAAAAAATAAAGACGTCGGTCTTTGTTGCAGATTGTTGACAA
S  A  T  G  I  L  I  K  N  K  D  V  G  L  C  C  R  L  L  T

      5470       5480       5490       5500       5510       5520
ACCTATTTATTGAACGAGAACGGTAGTAGCGGAATAGAAGAGTAAAGCGTTTGCGCCAGG
N  L  F  I  E  R  E  R  *  *

      5530       5540       5550       5560       5570       5580
GATGGCGCAATCCGAGCGTACAGGGATGTATTTACAGCGTCTTTACGATCTATCCGTTAC

      5590       5600       5610       5620       5630       5640
TACCGCCATACGAACTTTGTCGGCAACCTCAAATAACTATGCCAGTCTTTATTGGTGACT
                                ClaI
      5650       5660       5670
GTTGCAGATGGATTGTGGGTAGTTGCTATCGAT
```

6.4.2. Identification of open reading frames (orfs)

Four orfs were identified in this 5.7 kb region. These are shown in Figure 6.2. and have been named outD, outE, outF and outG. Potential translation initiation codons are indicated by a bold amino acid (M). Putative ribosome binding sites (RBS) have been underlined. Translation termination codons are illustrated with an *. No putative start sites (promoters) for transcription were detected. The presence of significant DNA repeats or hairpin forming sequences was not detected using the 'Microgenie' analysis function.

6.4.3. Molecular organisation of the four out genes

The four out genes were predicted to be expressed in the direction as indicated in Figure 6.2. There were no intergenic regions of DNA between the predicted coding regions of outD, outE and outF. The termination codons of outD and outE are predicted to overlap with the initiation codons of outE and outF respectively. A small intergenic region between outF and outG of 32 bases was predicted. The close juxtaposition of the genes, their unidirectional expression, and the potential for translational coupling suggests that these genes might be expressed as part of an operon.

Table 6.2. Amino acid content of Out proteins

OutD

The sequence contains 654 amino acids:

Ala	54	(8.3)	Leu	53	(8.1)
Arg	39	(6.0)	Lys	29	(4.4)
Asn	43	(6.6)	Met	19	(2.9)
Asp	40	(6.1)	Phe	21	(3.2)
Cys	1	(0.2)	Pro	21	(3.2)
Gln	33	(5.0)	Ser	56	(8.6)
Glu	31	(4.7)	Thr	47	(7.2)
Gly	46	(7.0)	Trp	6	(0.9)
His	3	(0.5)	Tyr	12	(1.8)
Ile	35	(5.4)	Val	65	(9.9)
End	1	(0.2)			

Acidic	(Asp + Glu)	71	(10.9)
Basic	(Arg + Lys)	68	(10.4)
Aromatic	(Phe + Trp + Tyr)	39	(6.0)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	211	(32.3)

Molecular Weight = 71279. (Immature) Average pI = 6.08
Molecular Weight = 68700. (Mature)

OutE

The sequence contains 499 amino acids:

Ala	38	(7.6)	Leu	66	(13.2)
Arg	45	(9.0)	Lys	12	(2.4)
Asn	14	(2.8)	Met	13	(2.6)
Asp	33	(6.6)	Phe	7	(1.4)
Cys	6	(1.2)	Pro	20	(4.0)
Gln	25	(5.0)	Ser	31	(6.2)
Glu	36	(7.2)	Thr	28	(5.6)
Gly	32	(6.4)	Trp	1	(0.2)
His	8	(1.6)	Tyr	9	(1.8)
Ile	38	(7.6)	Val	36	(7.2)
End	1	(0.2)			

Acidic	(Asp + Glu)	69	(13.8)
Basic	(Arg + Lys)	57	(11.4)
Aromatic	(Phe + Trp + Tyr)	17	(3.4)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	170	(34.1)

Molecular Weight = 55281. Average pI = 6.11

Table 6.2. (cont.)

OutF

The sequence contains 409 amino acids:

Ala	48	(11.7)	Leu	60	(14.7)
Arg	36	(8.8)	Lys	12	(2.9)
Asn	7	(1.7)	Met	23	(5.6)
Asp	16	(3.9)	Phe	11	(2.7)
Cys	4	(1.0)	Pro	14	(3.4)
Gln	25	(6.1)	Ser	34	(8.3)
Glu	19	(4.6)	Thr	16	(3.9)
Gly	20	(4.9)	Trp	1	(0.2)
His	10	(2.4)	Tyr	7	(1.7)
Ile	14	(3.4)	Val	31	(7.6)
End	1	(0.2)			

Acidic (Asp + Glu)	35	(8.6)
Basic (Arg + Lys)	48	(11.7)
Aromatic (Phe + Trp + Tyr)	19	(4.6)
Hydrophobic (Aromatic + Ile + Leu + Met + Val)	147	(35.9)

Molecular Weight = 45167. Average pI = 6.27

OutG

The sequence contains 157 amino acids:

Ala	6	(3.8)	Leu	17	(10.8)
Arg	8	(5.1)	Lys	9	(5.7)
Asn	9	(5.7)	Met	5	(3.2)
Asp	12	(7.6)	Phe	2	(1.3)
Cys	1	(0.6)	Pro	11	(7.0)
Gln	12	(7.6)	Ser	9	(5.7)
Glu	6	(3.8)	Thr	6	(3.8)
Gly	16	(10.2)	Trp	2	(1.3)
His	1	(0.6)	Tyr	7	(4.5)
Ile	7	(4.5)	Val	10	(6.4)
End	1	(0.6)			

Acidic (Asp + Glu)	18	(11.5)
Basic (Arg + Lys)	17	(10.8)
Aromatic (Phe + Trp + Tyr)	11	(7.0)
Hydrophobic (Aromatic + Ile + Leu + Met + Val)	50	(31.8)

Molecular Weight = 17353. Average pI = 6.05

6.5. Analysis of the predicted protein sequences encoded by the four out genes

6.5.1. Analysis of the primary structures of OutD, OutE, OutF and OutG using the 'Microgenie' package

6.5.1.1. Introduction

The 'Microgenie' protein analysis package provides an insight into the properties of the primary structures of both DNA and protein molecules. The Microgenie analysis package was used to predict the molecular weight of and determine the amino acid residue frequency for the OutD, OutE, OutF and OutG.

6.5.1.2. Results and discussion

The predicted primary structures of the Out proteins (D, E, F and G) are given in Figure 6.2. alongside the coding region of DNA. The predicted molecular weights and the amino acid residue frequencies of each Out protein are shown in Table 6.2.

OutD has two potential translation initiation codons. The first nucleotide of each codon is 241 and 256. The predicted sizes of the potential OutD protein are 654 amino acids and M_w 71.279 kD (initiation codon at nucleotide 241), or 649 amino acids and M_w 70.652 kD (initiation codon at nucleotide 256). Although the initiation codon at nucleotide position 256 is preceded by a good ribosome binding site (Gold and Stormo, 1987), the potential initiation codon at 241 is preferred for reasons given in section 6.5.2.2.1. A potential N-terminal signal-sequence was predicted for OutD (6.5.2.2.1.) which is indicated by double underlining in Figure 6.2. The

predicted signal-sequence processing site is indicated by an arrow.

The potential OutE protein is predicted to be composed of 499 amino acid residues and have an M_w of 55.281 kD. This protein is probably initiated by a codon starting at nucleotide position 2202. The putative ribosome binding site preceding the potential OutE initiation codon is underlined.

The predicted M_w of the OutF protein is 45.167 kD. This protein is predicted to comprise 409 amino acid residues. The potential initiation codon for OutF starts at nucleotide 3700. A potential ribosome binding site which precedes this codon is underlined.

OutG is predicted to have an M_w of 19.444 kD comprising 176 amino acid residues. The potential initiation codon for OutG starts at nucleotide 4952. A potential ribosome binding site for initiation at this codon is underlined.

6.5.2. Secondary structure predictions and feature searches for Out proteins

6.5.2.1. Introduction

The Out proteins were investigated using the Leeds University protein engineering software (LUPES) programme 'Matscan'. Matscan enables (probe) proteins to be screened against a database of short protein motifs (features). By finding homologies with short protein structures of known function, it is possible to gain an insight into properties of the protein under investigation. The University of Wisconsin genetics computer group (UWGCG) has developed a number of powerful biocomputing programmes (Devereux et al., 1984). The programme 'Signalpep' was used to identify the locations of potential signal-sequences in the Out proteins. The UWGCG programmes 'Proteinstructure' and 'Plotstructure' were used to plot the predicted secondary structures of

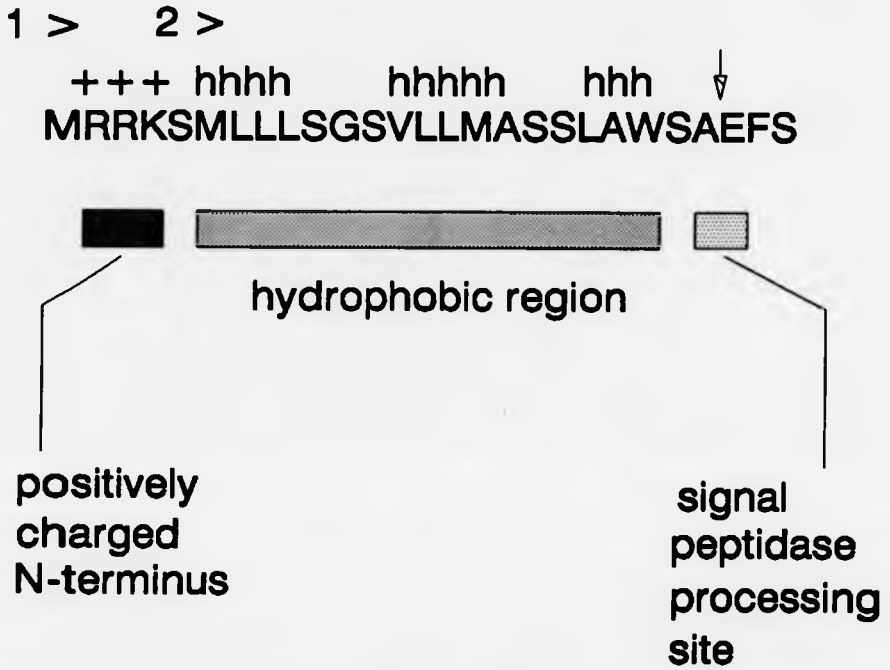
the Out proteins. The default algorithms of Chou and Fasman (1978) and Kyte and Doolittle (1982), which predict secondary structure and hydropathy respectively, were used. The predictions of 'Proteinstructure' are displayed as a 'squiggle plot' for each Out protein.

6.5.2.2. Results and discussion

6.5.2.2.1. OutD

The UWGCG 'Signalpep programme' identified a potential signal-sequence in OutD. The immature protein of M_w 71.279 kD (654 amino acid residues) is predicted to be processed to a mature form of M_w 68.700 kD (630 amino acid residues). The signal-sequence of the immature OutD protein contains 24 amino acid residues including an N-terminal positively charged region. However, a potential ribosome binding site upstream of the putative valine initiation codon (first nucleotide 256) was not found. An alternative potential translational start codon for this protein starts at nucleotide position 256. This initiation codon (methionine) is preceded by good potential 'ribosome binding site'. However, the N-terminal region of the resulting OutD pre-protein would not be positively charged. The features of the putative OutD signal-sequence which initiates at a codon starting at nucleotide 241 are shown in Figure 6.3. The positively charged 5' terminal region, an hydrophobic core and a typical cleavage site are illustrated in this figure (von Heijne, 1987). The hydrophobicity of the N-terminal signal-sequence is also illustrated in the 'squiggle plot' of OutD. (Figure 6.4.). No features were found for OutD using the Matscan programme of LUPES. The presence of this predicted signal-sequence suggests that OutD is probably targeted to the periplasm or beyond. This prediction is supported by the fact

Figure 6.3. OutD has a potential N-terminal signal-sequence



Key

⇩ predicted point of cleavage
\+ positively charged residue
h hydrophobic residue
> potential translation start

Figure 6.4 Secondary structure predictions for OutD

Key to symbols

α- helix

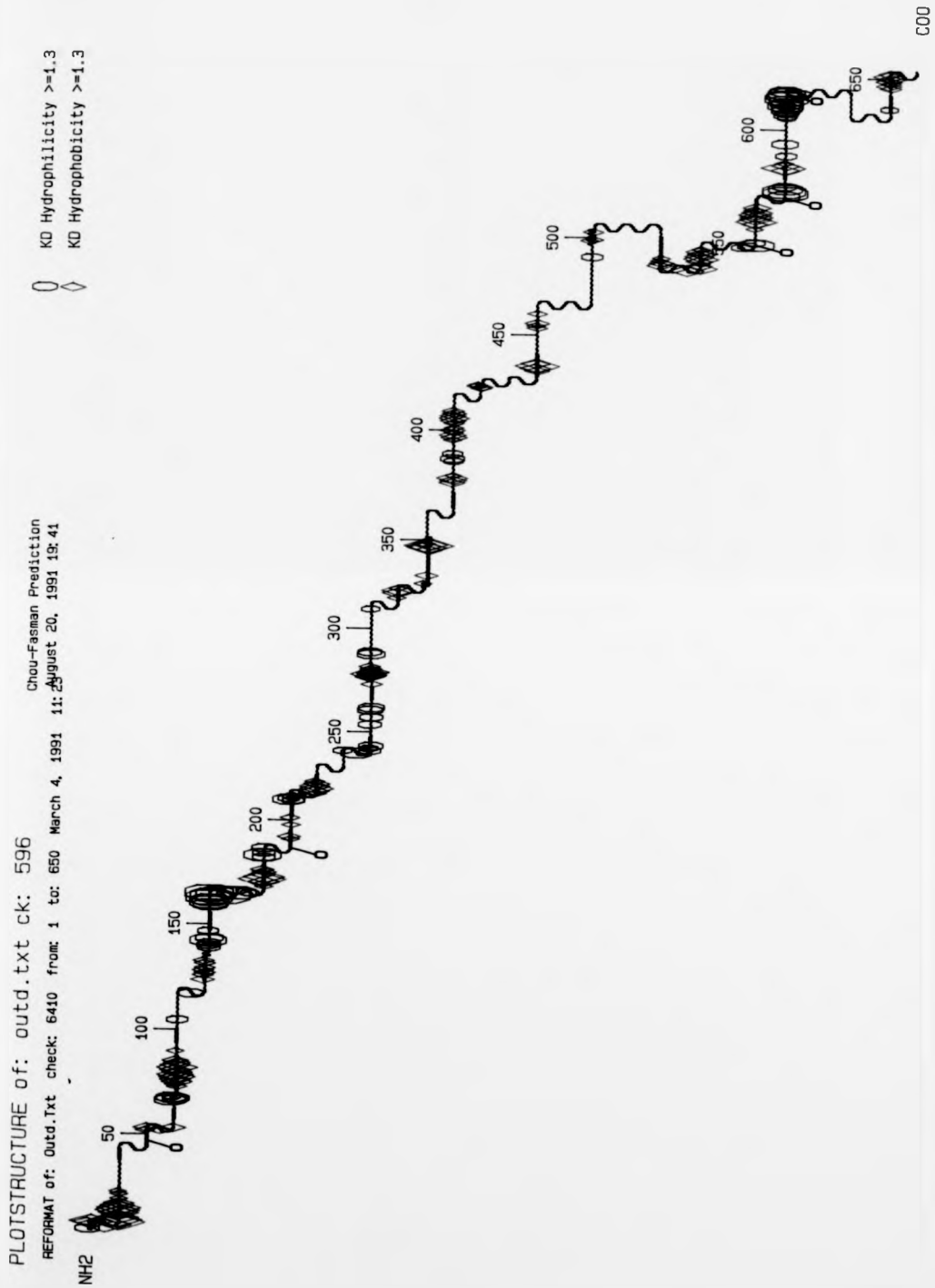
β- pleated sheet

β- turn region

random coil

Figure 6.4 (opposite page) shows a plot of predicted secondary structure for OutD using the algorithms of Chou and Fasman (1978). This 'squiggle plot' was achieved using the UWGCG programmes 'Proteinstructure' and Plotstructure (Devereux et al., 1984). A plot of hydrophobicity is superimposed onto this secondary structure plot using the algorithms of Kyte and Doolittle (1982) and a window of 10 residues. Regions of hydrophobicity are shown by green diamonds and regions of hydrophilicity are shown by red ovals.

OutD has a potential N-terminal signal-sequence which is illustrated by the high hydrophobicity at the extreme N-terminal region. This protein contains a large number of regions which are predicted to form β - sheet structures.



that a 'blue' TnphoA mutant (PR54) has been mapped (by sequence analysis) to outD which encodes OutD (M. Gibson, pers. comm.). This indicates that OutD is either periplasmic or has domains protruding into the periplasm from either the IM or the OM. The secondary structure predictions (Figure 6.4.) reveal a high incidence of β-sheet structures. Amphipathic β-sheet structures are a feature of multi-spanning integral OM proteins (Struyve et al., 1991; von Heijne, 1987). This suggests that OutD might be an OM protein. The level of predicted α-helix forming regions in OutD was low. The OutD protein does not have a C-terminal phenylalanine residue unlike a family of OM proteins recently described by Struyve et al. (1991).

6.5.2.2.2. OutE

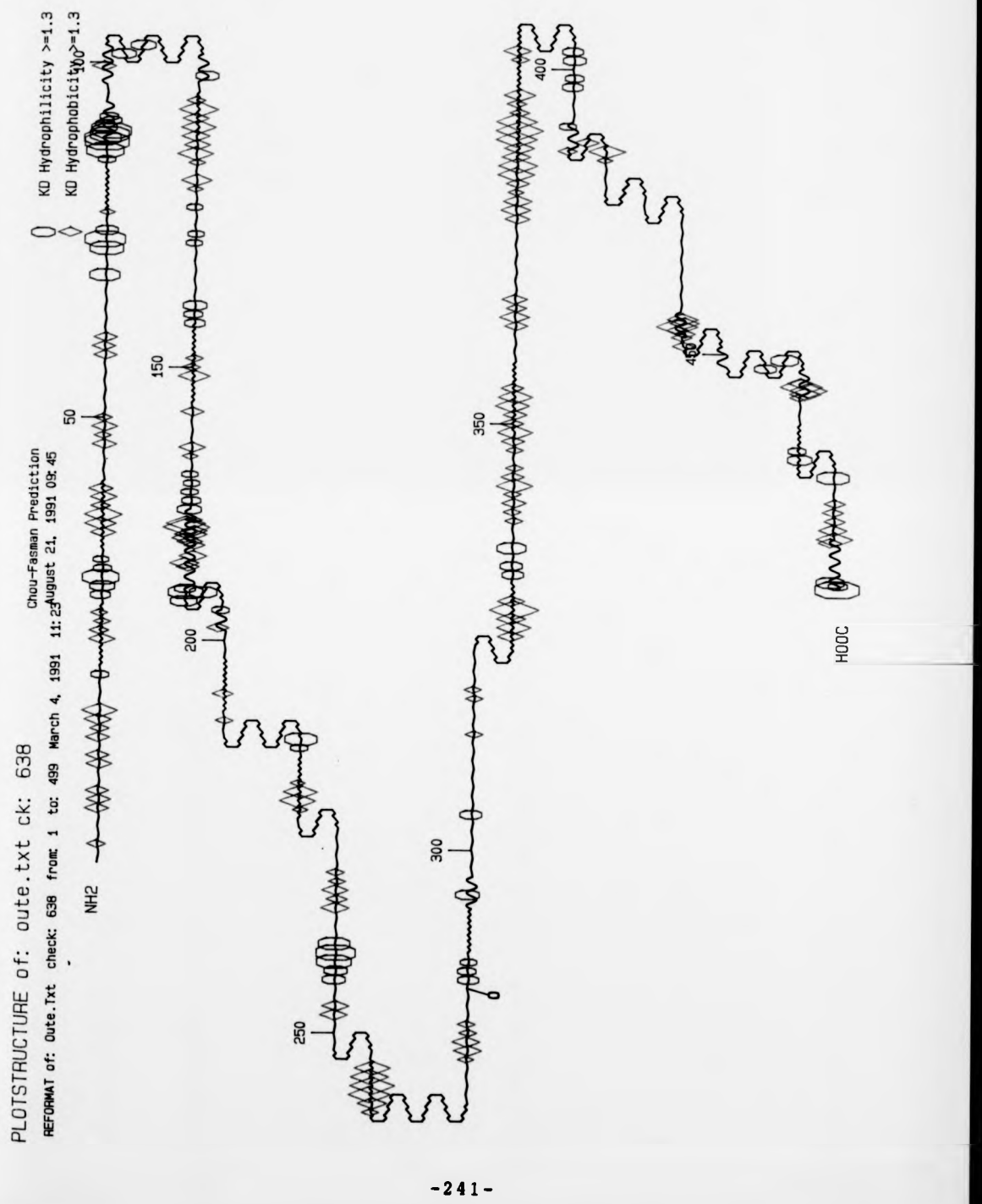
The LUPES Matscan programme identified two potential nucleotide binding site motifs in OutE. The first nucleotide binding site (Type A) has a consensus sequence G-X-X-X-X-G-K-T-X-X-X-X-X-Hy-Hy (where X represents any residue and Hy represents an hydrophobic residue) (Walker et al., 1982). This location of this putative nucleotide binding site is indicated in Figure 6.11. Another less strongly conserved putative nucleotide binding site (Type B), also shown in Figure 6.11., has a consensus sequence of R/K-X-X-X-G-X-X-X-L-Hy-Hy-Hy-Hy-Hy-D (Walker et al., 1982). The predicted OutE protein contains no potential signal-sequences or potential membrane-spanning regions (hydrophobic α-helices or extensive β-sheet structures) as demonstrated by the plot of secondary structure (Figure 6.5.). For these reasons, OutE is probably a cytoplasmic protein. This result is intriguing because it was initially expected that proteins involved in secretion (across the OM) would be associated with the IM, OM, or periplasm.

Figure 6.5. Secondary structure predictions for OutE

The key for the secondary structures is given in Figure 6.4.

Figure 6.5 (opposite page) shows a plot of predicted secondary structure for OutE using the algorithms of Chou and Fasman (1978). This 'squiggle plot' was plotted using the UWGCG programmes 'Proteinstructure' and Plotstructure (Devereux et al., 1984). A plot of hydrophobicity is superimposed onto this secondary structure plot using the algorithms of Kyte and Doolittle (1982) and a window of 7 residues. Regions of hydrophobicity are shown by green diamonds and regions of hydrophilicity are shown by red ovals.

No strongly hydrophobic membrane spanning regions were predicted for OutE. However, a long stretch of predominantly hydrophobic residues is present between amino acid residues 340 - 390. OutE is likely to be a soluble cytoplasmic protein.



6.5.2.2.3. OutF

OutF was analysed using the Signalpep programme of UWGCG but no potential signal-sequences were predicted. The LUPES programme 'Matscan' did not find any features in this protein. The secondary structure prediction plot, using the UWGCG programme 'Proteinstructure', did reveal some interesting characteristics of OutF. The predicted secondary structure of OutF is presented in Figure 6.6. OutF is predicted to contain strongly hydrophobic regions. Two hydrophobic α-helix forming regions are predicted at amino acid positions 90 to 100 and 240 to 250. A small region of hydrophobic α-helix is predicted at 280 to 290. The most intensely hydrophobic regions are predicted to form β-sheet structures. Two such regions are located at amino acid positions 170 to 200 and 370 to 409. These regions are possibly membrane anchors. Membrane anchors are a characteristic feature of IM proteins (von Heijne, 1987). This data suggests that OutF might be located in the IM and might span this membrane several times.

6.5.2.2.4. OutG

No features were predicted for OutG by the LUPES or UWGCG packages. The plot of secondary structure (Figure 6.7.) shows that OutG is predicted to contain one highly hydrophobic N-terminal β-sheet forming domain. This is located between amino acid residues 20 and 40 and is suggestive of a potential signal-sequence. However, this N-terminal region was not predicted to be a classical signal-sequence by the 'Signalpep' programme of the UWGCG package. This hydrophobic region might be a membrane spanning domain. It was not until OutG was screened against a protein data bank that the significance of this N-terminal hydrophobic region

Figure 6.6. Secondary structure predictions for OutF

The key for the secondary structures is given in Figure 6.4.

Figure 6.6 (opposite page) shows a plot of predicted secondary structure for OutE using the algorithms of Chou and Fasman (1978). This 'squiggle plot' was plotted using the UWGCG programmes 'Proteinstructure' and Plotstructure (Devereux et al., 1984). A plot of hydrophobicity is superimposed onto this secondary structure plot using the algorithms of Kyte and Doolittle (1982) and a window of 20 residues. Regions of hydrophobicity are shown by green diamonds and regions of hydrophilicity are shown by red ovals.

OutF is predicted to contain some strongly hydrophobic domains. These are indicated by the intensity and size of the green diamonds in this figure. The major areas of hydrophobicity coincide with regions which are predicted to form strong secondary structures. These are α-helix forming regions (residues 90 - 100 and 240 - 250) and β-sheet forming regions (residues 170 - 200 and 370 - 409).

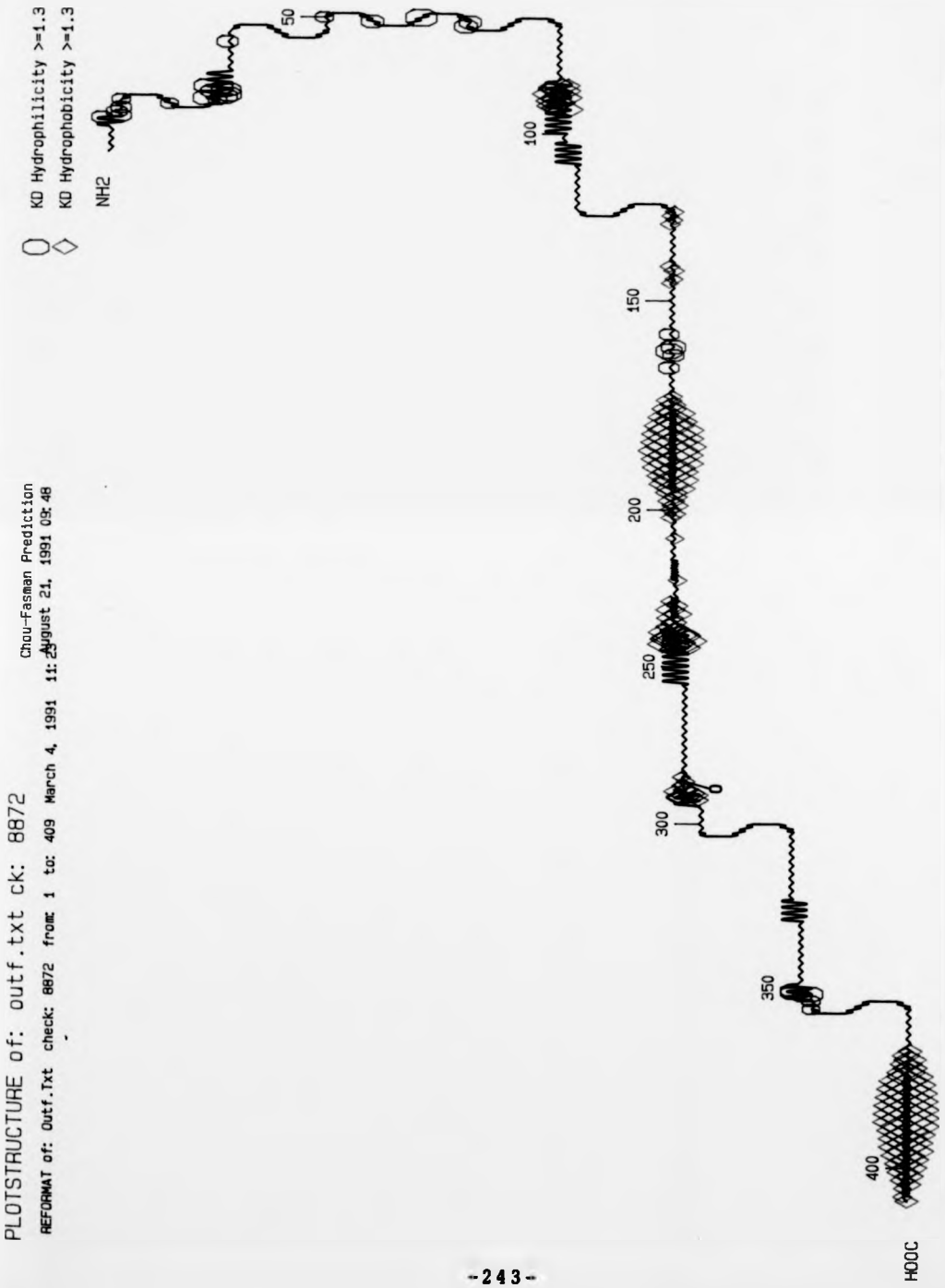
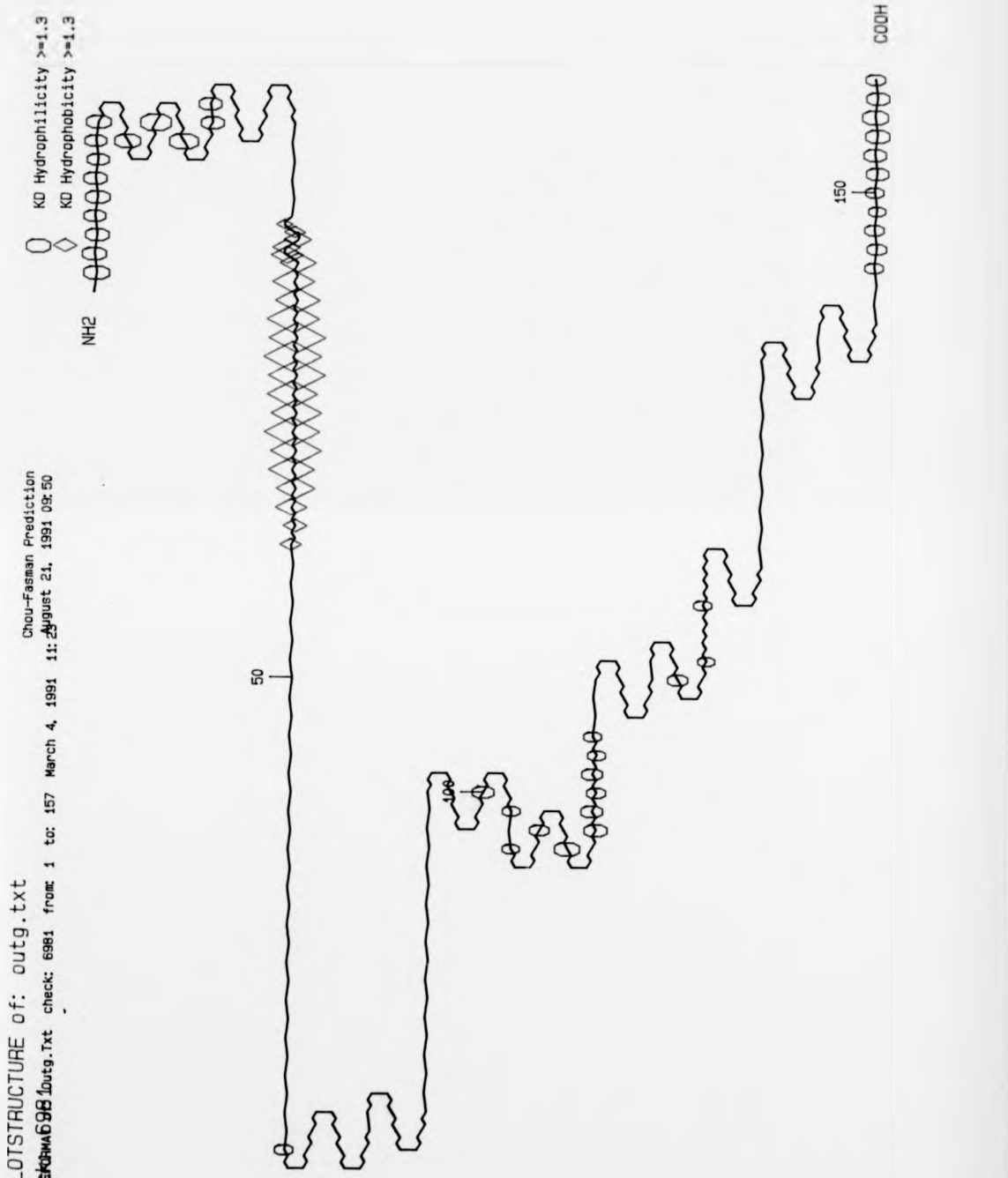


Figure 6.7. Secondary structure predictions for OutG

The key for the secondary structures is given in Figure 6.4.

Figure 6.7 (opposite page) shows a plot of predicted secondary structure for OutE using the algorithms of Chou and Fasman (1978). This 'squiggle plot' was plotted using the UWGCG programmes 'Proteinstructure' and Plotstructure (Devereux et al., 1984). A plot of hydrophobicity is superimposed onto this secondary structure plot using the algorithms of Kyte and Doolittle (1982) and a window of 20 residues. Regions of hydrophobicity are shown by green diamonds and regions of hydrophilicity are shown by red ovals.

One hydrophobic β-sheet forming region was predicted for OutG at residues 20 - 40 which might be a membrane spanning region of this protein.



was realised. This feature will be discussed in section 6.8.5. The OutG protein has no hydrophobic regions (>=1.3) beyond amino acid position 40. This suggests that it probably spans the IM once.

6.6. Homologies between the Out proteins and proteins contained in the OWL protein data bank

6.6.1. Introduction

The Out proteins were used as probes to search a protein database. The aim of this search was to find proteins that shared sequence homologies with the Out proteins. By investigating the functions and properties of any such proteins, especially those with a known function, it should be possible to obtain more information about the possible functioning of the Out proteins. The protein data bank used was OWL 3.02 (1991). The OWL database comprises seven general protein databases, these being NBRF-PIR, SWISSPROT, GENBANK, NBRF-PIR NEW, NEWAT86 PSD-KYOTO and BROOKHAVEN. The OWL database was searched using NEWSWEEP, a programme of the Leeds University protein engineering software (LUPES).

Predicted protein sequences of molecular trafficking systems from other Gram-negative bacteria were also obtained. These were unpublished at the time of this analysis and included Xps proteins from Xanthomonas campestris (M. Dow, pers. comm.) and Xcp from Pseudomonas aeruginosa (A. Lazdunski, pers. comm.). The xps DNA and Xps protein sequence data have recently been published (Dums et al., 1991). Dums et al. (1991) sequenced a 5 kb region of DNA from a cluster of DNA which encodes secretion and pathogenicity functions. We were provided with proteins encoded by xpsE (XpsE), xpsF XpsF), xpsG (XpsG), xpsH (XpsH) and xpsI

(Xpsl). The DNA sequence obtained from the laboratory of A. Lazdunski was from the 0' and 55' regions of the P. aeruginosa chromosome. The 55' region contains xcp-5, xcp-51 and xcp-52 and the 0' region contains xcp-1 (xcpA)(Lazdunski et al., 1990). We were provided with the predicted protein sequences encoded by xcp-1 (XcpA), xcp-51 (Xcp-51) and xcp-52 (Xcp-52). Other xcp loci have been identified at 62' (xcp-6) and at 55' (xcp-53, xcp-54 and xcp-55) on the P. aeruginosa chromosome (Lazdunski et al. 1990). The xcp genes are involved in the secretion of lipase, elastase, alkaline phosphatase, phospholipase C and exotoxin A from P. aeruginosa (Lazdunski et al., 1990).

6.6.2. Results

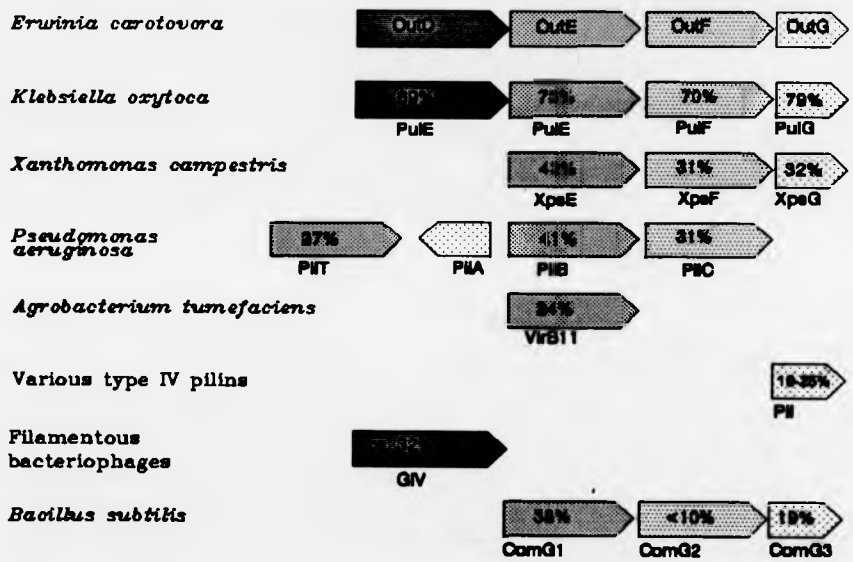
The four Out proteins were screened against a protein database in order to identify proteins sharing amino acid sequence similarities. Significant similarities were found between the Out proteins and proteins involved in the secretion of extracellular enzymes from other Gram-negative bacteria. These included Pul proteins encoded by the pulC - pulO operon (involved in the secretion of pullulanase from Klebsiella oxytoca) and Xps proteins (involved in the secretion of pectinases, cellulases and amylases from Xanthomonas campestris). The predicted Xcp proteins from P. aeruginosa were not similar to OutD, OutE, OutF or OutG. However, when searched against the Pul proteins the following was found. Xcp-51 was similar to PulL (34%), Xcp-52 was similar to PulM (25%) and XcpA (encoded by xcp-1) was similar to PulO (38%). Protein similarity was found between OutE and PilB and also between OutF and PilC. PilB and PilC are involved in the biogenesis of type IV (MePhe) pili (Nunn et al., 1990). Even more intriguing is the similarity between OutG and pilin monomers from a number of different bacterial

species (Figure 6.8). Similarity was also discovered between the Out proteins and macromolecular transport systems found in the Gram-positive bacterium, Bacillus subtilis (ComG3 - involved in DNA uptake) and Agrobacterium tumefaciens (VirB11 - involved in the transfer of the Ti plasmid DNA out of the bacterial cell). Similarity between OutD and the geneIV protein product (pIV) of a number of filamentous bacteriophages (fd, f1, m13, Ike) was also identified. These findings are summarised in Figure 6.8. and will be discussed in detail in the following sections. The percentages show the proportion of identical residues which are conserved in the Out proteins and their homologues. As well as significant identities between the actual proteins of these different bacteria, there are strong similarities in the genetic organisation of the out, pul and xps genes. The organisation of pilB and pilC is also conserved. However, PilD is not homologous to OutG but interestingly is homologous to the last protein in the PulC-O operon, PulO. In a separate piece of work we have recently demonstrated that OutO from Ecc is also homologous to PilD. This work will not be presented in this thesis.

In all four groups, protein similarities were identified which will be discussed in greater detail in the following section.

The first group contains OutD, PulD and pIV from a variety of filamentous bacteriophages (Pf3 [p430], fd, f1, M13, Ike). The second group comprises OutE, PulE, XpsE, VirB11, PilB, PilT and ComG. The third group comprises OutF, PulF, XpsF and PilC. The final group comprises OutG, PulG and XpsG and type IV pilins from a variety of bacteria. The protein nomenclature for these proteins is described in the figure legends of Figures 6.10., 6.12., 6.14. and 6.16. The sources of these proteins and their biological roles will be discussed in the following sections.

Figure 6.8. Homologous membrane traffic proteins



6.7. Multiple alignment studies with Out and related proteins

6.7.1. Introduction

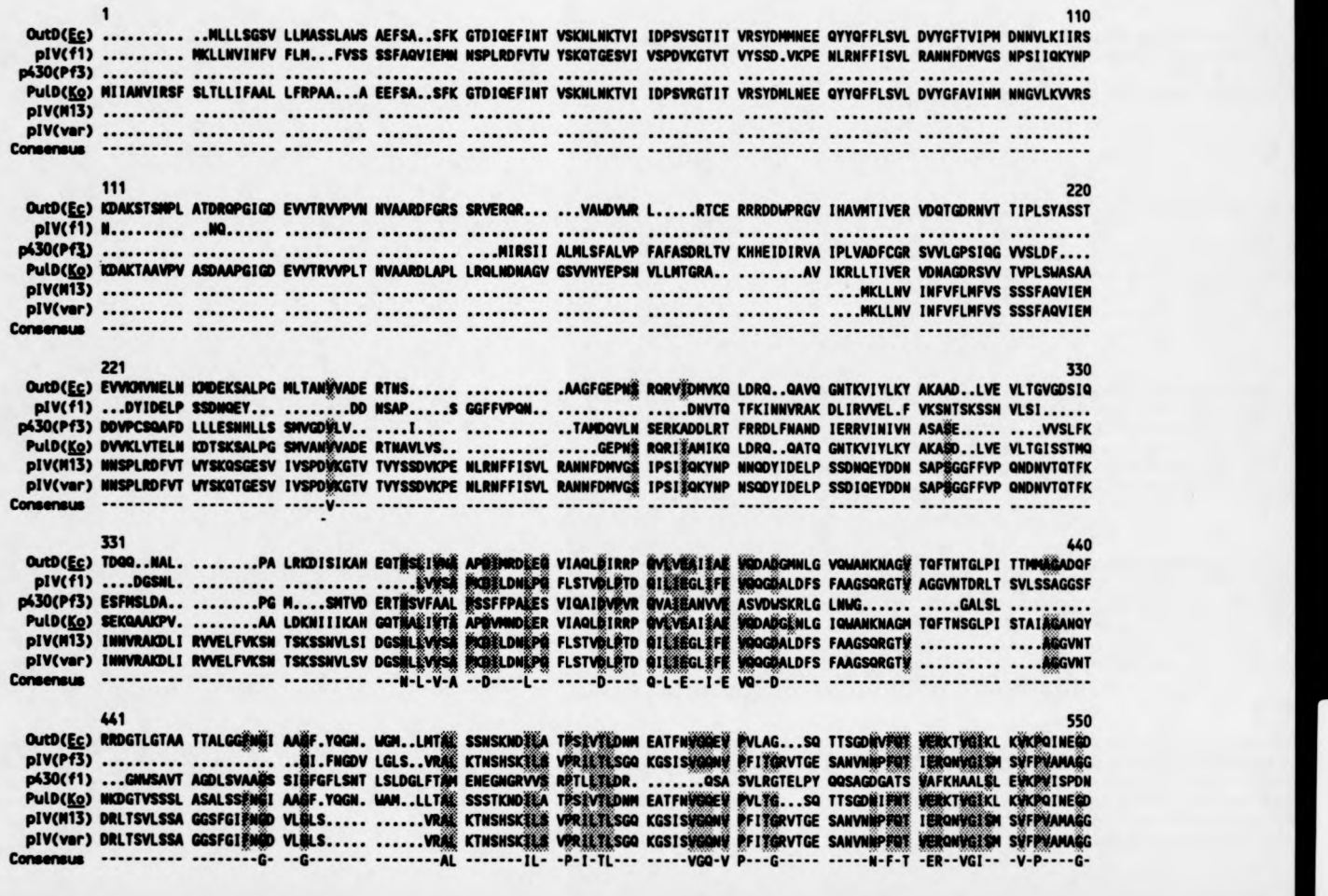
In order to identify areas of similarity between the Out proteins and their homologues, multiple protein alignments were performed using a programme called 'Gapzero'. Another programme, 'Root66', which analyses the degree of relatedness (shown as a dendrogram) between groups of proteins was also used. Both 'Gapzero' and 'Root66' are programmes obtained from SEQUNET (SERC Daresbury Laboratory) which were written by Julian Parkhill and Duncan Rouch. For OutG, a variety of pilins from different bacterial species were chosen. Several pIV proteins from various filamentous bacteriophages were analysed alongside the OutD protein. The results are presented separately for each Out protein.

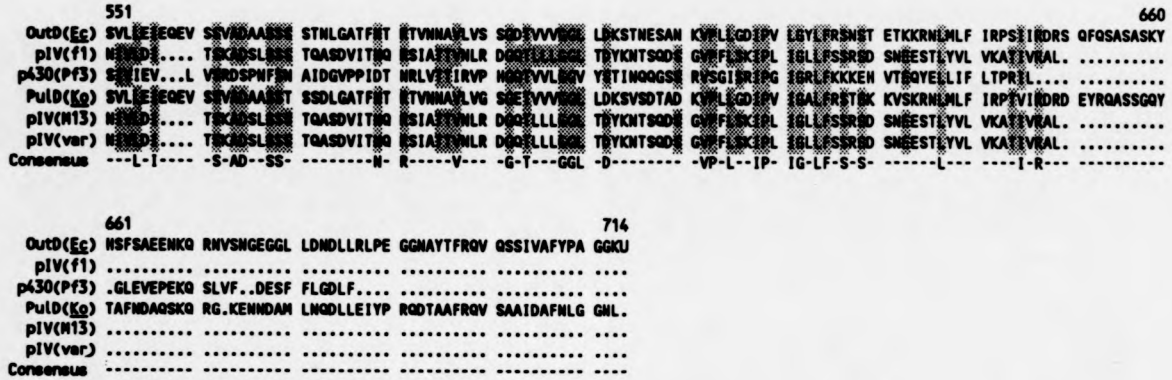
6.7.2. Results and discussion

6.7.2.1. OutD homologues: alignment of protein sequences

A multiple alignment of the OutD homologues is shown in Figure 6.9. A dendrogram showing the relationship between the OutD homologues is given in Figure 6.10. The figure legend succeeding Figure 6.10. shows the proteins used in both Figure 6.9. and Figure 6.10. The OutD protein shares similarity across its entire sequence with PulD and the pIV proteins. The dendrogram reveals that OutD and PulD are more closely related to each other than they are to any of the pIV proteins.

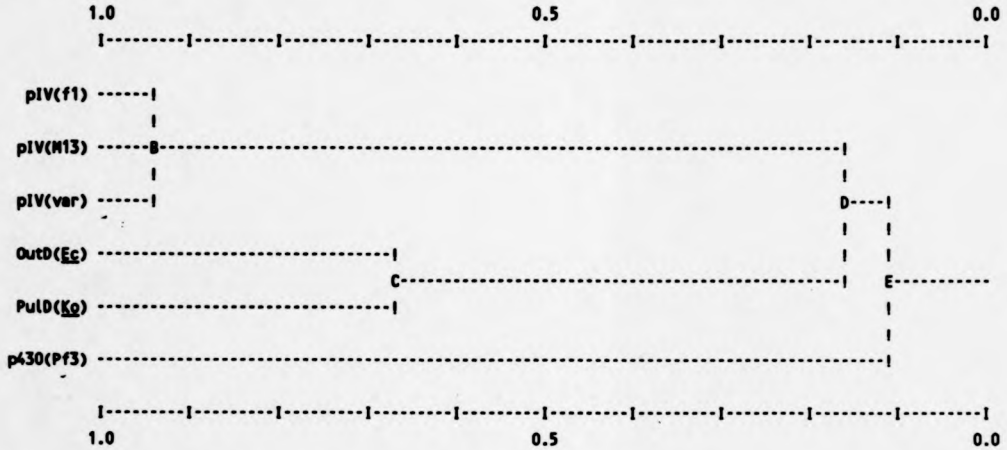
Figure 6.9. Multiple alignment of proteins related to OutD





Positions where 4 out of 6 proteins are identical are shaded. The consensus shows positions where 5 from 6 proteins are identical. The protein symbols are explained in the legend which follows Figure 6.10.

Figure 6.10. Dendrogram showing the relationship between OutD and homologous proteins



Legend for Figure 6.9. and Figure 6.10.

The predicted proteins used in Figure 6.9. and 6.10. are as follows: OutD protein from *Erwinia carotovora* (Ec) is from this work; GeneIV(pIV) proteins from various filamentous bacteriophages [f1 (Beck and Zinc, 1981), fd (Beck et al., 1978), M13 (van Wezenbeek *et al.*, 1980)]; pIV from f1 (Beck and Zinc, 1981), pIV(m13) from M13 (van Wezenbeek *et al.*,1980); p430 from the *Pseudomonas aeruginosa* filamentous bacteriophage Pf3 (Luiten *et al.*, 1985) and PulD from *Klebsiella oxytoca* (Ko) (d'Enfert *et al.*, 1989).

6.7.2.2. OutE homologues: alignment of protein sequences

The alignment of the OutE homologues is shown in Figure 6.11. A dendrogram showing the relationship between OutE homologous proteins is shown in Figure 6.12. The proteins used in both figures are presented in the legend for Figure 6.12. OutE, PulE, XcpE, VirB11, ComG3 and PilB exhibit homology across their entire sequences. PulE is a another protein of the pulC-O operon from K. oxytoca (Pugsley et al., 1990a). The homology between OutE and XpsE is lower than that between OutE and the K. oxytoca homologue (PulE). However, this finding is not surprising as Ecc and X. campestris are not as closely linked taxonomically as are Ecc and K. oxytoca. X. campestris is more closely related to the pseudomonad family of bacteria, the DNA of which is high in G/C nucleotides.

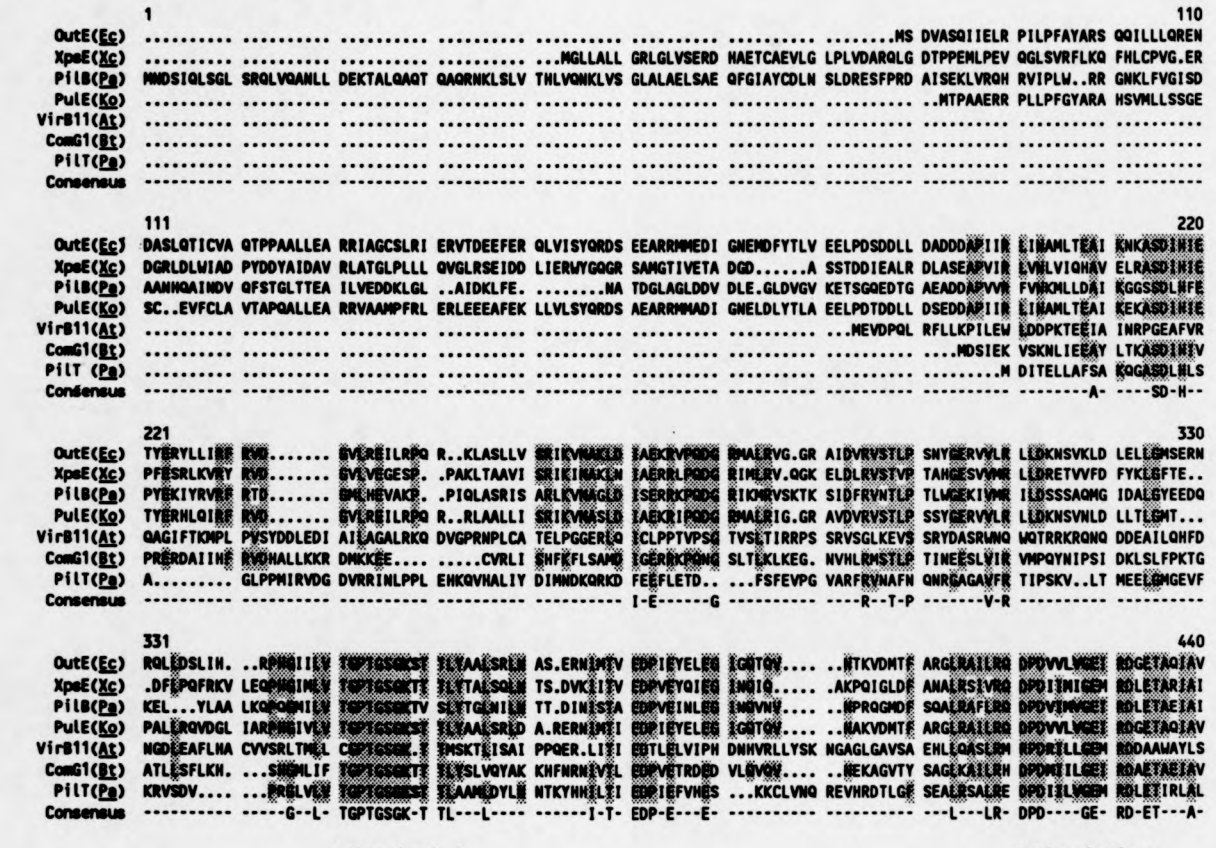
The PilB protein is encoded by the pilB gene of Pseudomonas aeruginosa and is one of several proteins which are required for the assembly of type IV (MePhe) pili in this bacterium (Nunn et al., 1990).

The homology between VirB11 and ComG1 has been reported previously (Albano et al., 1989). Particularly striking was the presence of potential nucleotide binding sites in all the OutE homologues. These sequences were found to be strongly conserved in all of these proteins.

6.7.2.3. OutF homologues: alignment of protein sequences

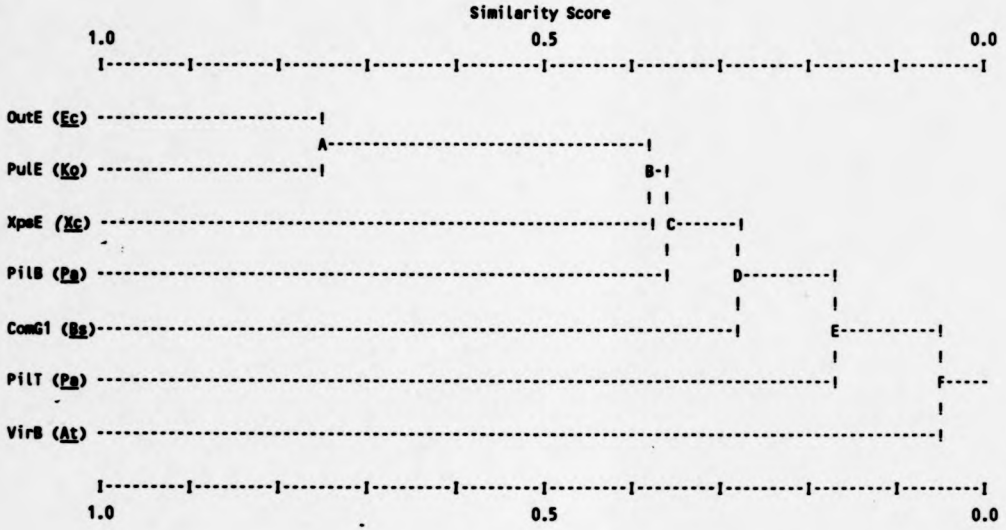
The alignment of OutF and related proteins is given in Figure 6.13. A dendrogram which shows the relationship between the OutF family of proteins is given in Figure 6.14. The alignment of this group of proteins shows that they are similar throughout their entire sequences. The homology was strongest between OutF and PulF, a feature clearly illustrated by the dendrogram.

Figure 6.11. Multiple alignment of proteins related to OutE



Positions where 4 out of 7 proteins are identical are shaded. The consensus shows positions where 5 out of 7 proteins are identical. Two potental nucleotide binding regions are in a bold type face and are discussed in the text. The protein symbols and their source are described in the legend following Figure 6.12

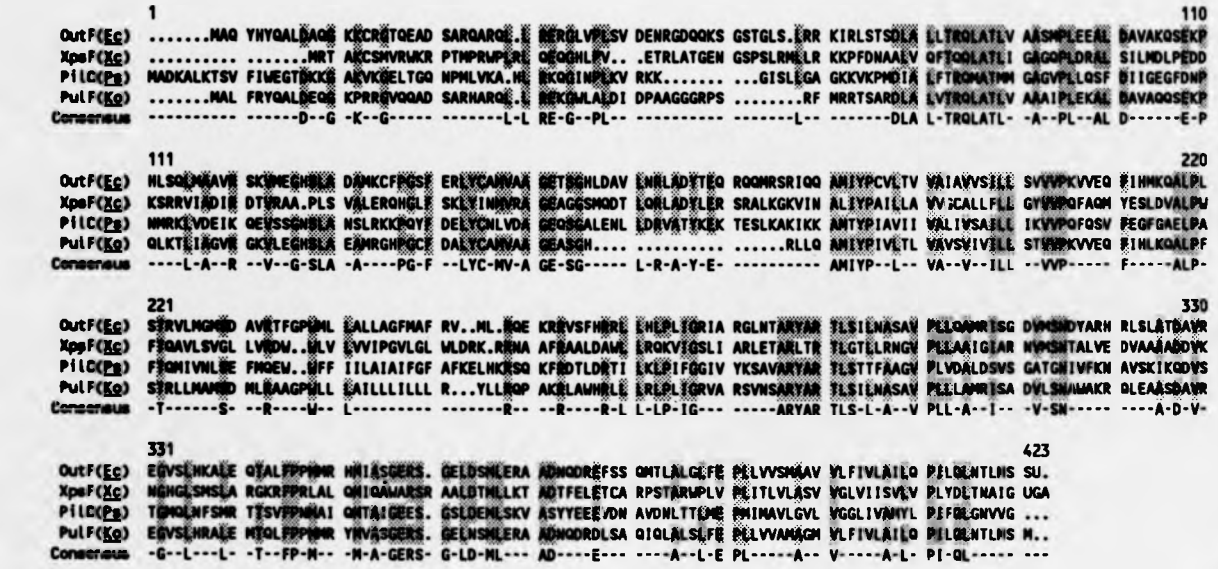
Figure 6.12. Dendrogram showing the relationship between OutE and homologous proteins



Legend for Figure 6.11. and 6.12.

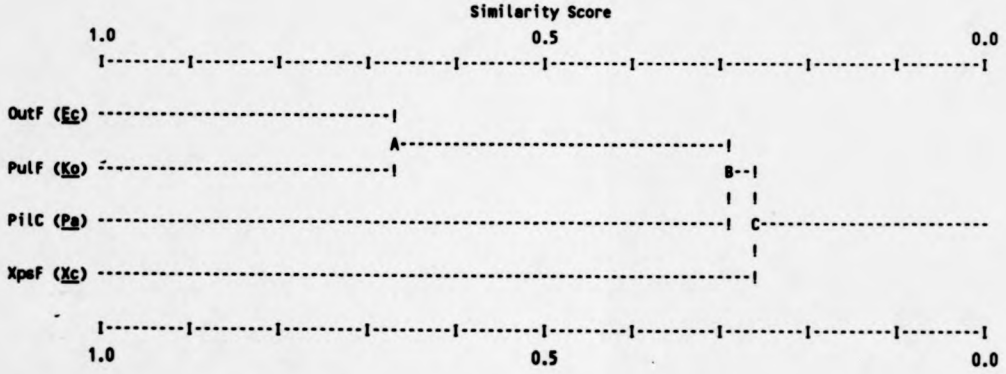
The predicted proteins used in Figure 6.11. and 6.12. are as follows: OutE from *Erwinia carotovora* (Ec)is decribed in this work; PulE from *Klebsiella oxytoca* (Ko) has not been published but was obtained from the Swissprot entry PulE$Klepn; XpsE from *Xanthomonas campestris* (Xc) (Dums *et al.*, 1991 in press); PilB (Nunn *et al.*, 1990) and PilT (Whitchurch *et al.*, 1990) from *Pseudomonas aeruginosa* (Pa) ; ComG1 (ORF1 of the *comG* operon) from *Bacillus subtilis* (Ba) (Albano *et al.*, 1989) and VirB11 encoded by ORF11 of the *virB* operon located on the Ti plasmid of *Agrobacterium tumefaciens* (At) (Ward *et al.*, 1988).

Figure 6.13. Multiple alignment of proteins related to OutF



Positions where 3 out of 4 proteins are identical are shaded. The consensus sequence also shows positions where 3 out of 4 proteins are identical. The protein symbols are explained in the legend which follows Figure 6.14.

Figure 6.14. Dendrogram showing the relationship between OutF and homologous proteins



Legend for Figure 6.13. and 6.14.

The predicted proteins used in Figure 6.13. and Figure 6.14. are as follows: OutF from *Erwinia carotovora* (Ec) is described in this work; PulF from *Klebsiella oxytoca* (Ko) has not been published but was obtained using the Swissprot entry PulE$Klepn; PilC from *Pseudomonas aeruginosa* (Pa) (Nunn *et al.*, 1990) and XpsF from *Xanthomonas campestris* (Xc) (Dums *et al.*, in press).

6.7.2.4. OutG homologues: alignment of protein sequences

The multiple alignment of OutG and related proteins is presented in Figure 6.15. The relationship between the OutG family of proteins is shown in Figure 6.16. The multiple alignment shows that homology extends throughout the entire length of OutG and PulG. XpsG is 32% similar (identity) to OutG. The highest similarity is between the OutG and PulG proteins. The similarity between OutG and the type IV pili from a variety of Gram-negative bacteria was limited to the N-terminal region. This is clearly demonstrated by the multiple alignment in Figure 6.15. The N-terminal region of type IV pilin monomers contains a processing site unique to this class of proteins (and related proteins). This processing site is called the MePhe cleavage site (Dalrymple and Mattick, 1987). Mature pilin monomers have six to eight residues removed from the immature form and the resulting N-terminal phenylalanine is methylated (Hermodson et al., 1978; Dalrymple and Mattick, 1987). The site of methylation and the first residue of mature pilin monomers is indicated by an * in Figure 6.15. XpsG and TcpA both have Met residues in place of Phe at the methylation site. It will be interesting to determine if this residue (Met) is processed and methylated as if it were a Phe residue.

6.8. Components of a specialised eubacterial macromolecular trafficking apparatus

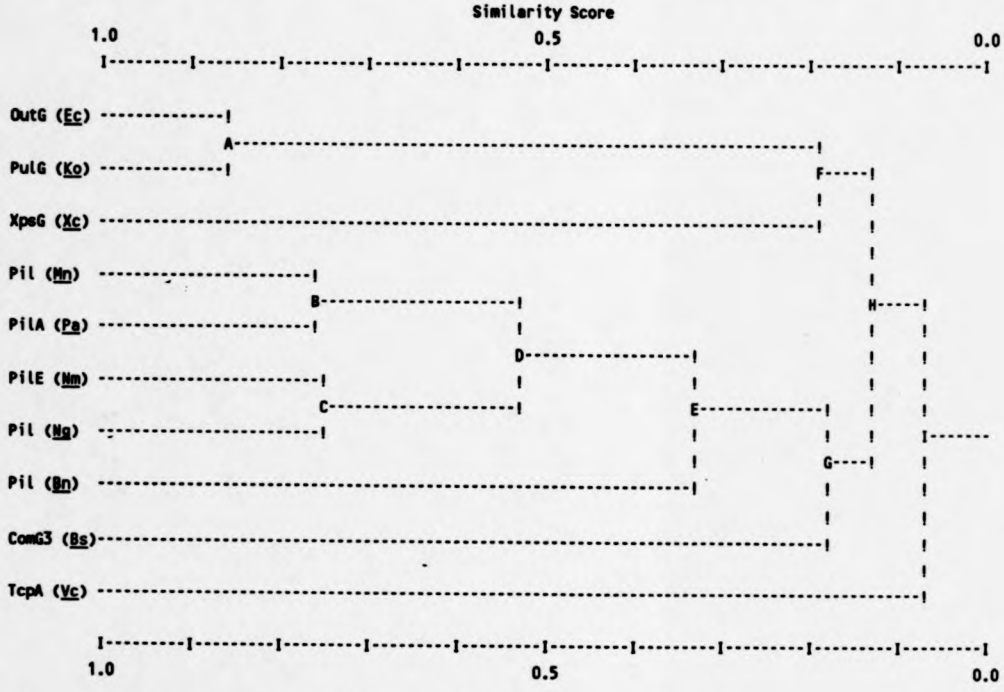
6.8.1. Introduction

The two-step secretion systems of Erwinia (Pel and Cel), K. oxytoca (pullulanase), P. aeruginosa (exotoxin A, phospholipase C, alkaline phosphatase, phospholipase, and elastase) and X. campestris (endoglucanases,

Figure 6.15. Multiple alignment of proteins related to OutG

Positions where at least 6 out of 10 proteins are conserved are shaded. The consensus shows positions where at least 7 out of 10 proteins are conserved. The N-terminal residue of the mature pilin monomer (MePhe) is marked with an asterisk (see text). The protein symbols are explained in the legend which follows Figure 6.16.

Figure 6.16. Dendrogram showing the relationship between OutG and homologous proteins



Legend for Figure 6.15. and Figure 6.16.

The predicted proteins used in Figure 6.15. and 6.16. are as follows: OutG from *Erwinia carotovora* (Ec) is described in this work; PulG from *Klebsiella oxytoca* (Ko) (Reyss and Pugsley, 1990); XpsG from *Xanthomonas campestris* (Xc) (Dums *et al.*, 1991); pilin monomers from *Moraxella nonliquefaciens* (Mn) (Froholm and Sletten, 1977), *Pseudomonas aeruginosa* (Pa) (Pasloske *et al.*, 1988), *Neisseria meningitidis* (Mn) (Potts and Saunder, 1988), *Neisseria gonorrhoeae* (Ng) (Meyer *et al.*, 1984), *Bacteroides nodosus* (Bn) (Finney *et al.*, 1988), *Vibrio cholerae* (Vc) (Faast *et al.*, 1989) and ComG3 (ORF3) from the *Bacillus subtilis* (Bs) *comG* operon (Albano et al., 1989).

pectinases and proteases) have been mentioned in section 1.7.6.). The secretion of pullulanase from K. oxytoca is the best characterised of these systems. Although mutants have been made which map to the out gene cluster, it is difficult to determine the role of each protein. An insight into the possible functioning of Out proteins might be gained by analysing the functions of similar proteins in other bacteria. In sections 6.8.2 to 6.8.5 the putative roles of OutD-G homologues in other transport systems will be discussed.

6.8.2. OutD homologues: functional roles

The role of OutD in the secretion of Pel and Cel from Ecc is not known. A TnphoA mutant (PR54) was isolated which mapped to this region, which might suggest that this gene is essential for secretion. However, the Out- phenotype of PR54 might be caused by a downstream polar effect of the transposon TnphoA. OutD, pIV and PulD proteins all have potential N-terminal signal-sequences. For PulD, protein sequencing experiments have confirmed that the mature OutD protein is probably processed at the predicted cleavage site (d'Enfert et al, 1989). Furthermore, PulD has been localised to the OM when expressed in K. oxytoca or E. coli. (d'Enfert et al., 1989). However, PulD was also found in the IM fraction. PulD is predicted to contain a high proportion of β-sheet forming regions. β-sheet structures are believed to form membrane spanning regions which are typical of OM proteins (Struyve et al., 1991). The role of the PulD protein is not known although it is essential for the secretion of pullulanase.

The morphogenesis of filamentous bacteriophages is well characterised (Russel, 1991; Kuhn et al., 1990a and b; Brissette and Russel, 1990; Shon et al., 1991; Nambudripad et al., 1991). The pIV protein of bacteriophages f1,

fd and M13 is thought to be involved in phage morphogenesis (Peeters et al., 1985; Beck and Zinc, 1991). It is also likely that the P430 protein of the P. aeruginosa filamentous bacteriophage performs a similar role (Luiten et al., 1985). pIV is not a component of the virion (Brissette and Russel, 1990), but might be part of a structure responsible for the release of the bacteriophage from the cell. In a review on filamentous phage assembly, Russel (1991) suggests that pIV might form a pore in the OM through which the assembled phage extrudes.

The cellular expression of pIV in E. coli induces a phage shock protein (Psp). The Psp protein is not required for filamentous bacteriophage production (Brissette et al., 1990). Psp is also induced under heat-shock, however the heat shock factor σ^{32} is not required. pIV must be correctly localised (beyond the cytoplasm) to induce this response (Brissette et al., 1990, 1991). The similarity between pIV and a protein from Haemophilus influenzae involved in DNA uptake has recently been reported (Russel, 1991), but this work is yet to be published.

6.8.3. OutE homologues: functional roles

The role of OutE in the secretion of Pel and Cel from Ecc is not known. However plasmid p3.7R1, which contained only one complete ORF (OutE), complemented Out- mutant, RJP122 and restored the Out+ phenotype (D. Whitcombe, pers. comm). This suggests that OutE is essential for the secretion of Pel and Cel from Ecc. The necessity of PulE for pullulanase exposition has not been proven (Pugsley et al., 1990a). XpsE is encoded by a region of DNA from X. campestris which is involved in the secretion of Pel, Cel and amylase from this phytopathogenic bacterium (Dow et al., 1987; Dums et al., 1991). However, the role of XpsE is not known.

The homology between OutE and both ComG1 and VirB11 is particularly exciting. ComG1 is encoded as part of the comG operon from Bacillus subtilis and is involved in the uptake of DNA (competence) in this bacterium (Albano et al., 1989). Genetic studies have identified several loci involved in competence in this bacterium. The comG operon has previously been referred to as the late competence locus (Albano et al., 1989). Transposon (Tn917lac) mutations in comG1 have been constructed. Strains carrying such mutations were deficient for competence. However this might have been due to the polarity of Tn917lac mutations on upstream genes of the comG operon (Albano et al., 1989).

The virB11 gene is part of the virB operon present in Agrobacterium tumefaciens (Ward et al., 1990; Shirasu et al., 1990; Kuldau et al., 1990; Christie et al., 1989). The virB locus is the largest of six distinct loci of vir genes and comprises 11 open reading frames (Kuldau et al., 1990). The virB operon is carried on the Ti plasmid and is involved in the transfer of the T DNA (Ti plasmid) from the bacterium to the plant host cell (Kuldau et al., 1990). The VirB11 protein has been over-expressed in E. coli and purified (Christie et al., 1989). Christie et al. (1989) demonstrated that VirB11 was able to hydrolyse ATP and had autophosphorylation activity in vitro. The ability of VirB11 to hydrolyse ATP is consistent with the predictions from the DNA sequence with respect to its potential nucleotide binding capabilities. It will be interesting to determine if the rest of the OutE family also have ATPase activity.

The PilC protein might be involved in the production of energy required for type IV pili assembly. Recently a gene was isolated which was found to complement the pil^{nr} mutants (Whitchurch et al., 1990). The pil^{nr} (non-retractile) mutations gave rise to a number of phenotypes including loss

of 'twitching' motility, hyper-piliation and resistance to certain bacteriophages (Bradley, 1980). The explanation for these observations was that the pili were no longer able to retract thus leading to hyper-piliation. It was postulated that the retraction of the pili might also be involved in bringing certain bacteriophages into close contact with the cell and thus aiding their entry. The pilT protein product (PilT) which complemented the Pil^{nr} phenotype exhibits sequence similarity with PilC (involved in pilus assembly). A potential nucleotide binding site was identified in PilT and it might be that these two proteins (PilC and PilT) act in concert to assemble and retract pili. Such a mechanism is perhaps the basis of motility provided by these pili (Whitchurch et al., 1990). It was mentioned in section 4.4. that a single Out- mutant (RJP190) was found to be resistant to two bacteriophages (ØD-2 and Ø565). The relevant mutation was found to map within the region which included outE (section 5.9.). It was not possible to further subclone this fragment and assign mutations to distinct reading frames. It is possible that the bacteriophage resistant phenotype of this mutant might be due a mutation in a gene which is responsible in energising the mechanism of entry of these two bacteriophages into the cell. However, preliminary studies suggest that this Out- mutant (RJP190) is unable to properly adsorb these two bacteriophages to its cell surface (section 4.4.). A possible explanation for this result is that a receptor molecule required for entry is lacking or unrecognisable. However, if a component of the Out apparatus was the receptor it might be expected that mutations in any of the genes required for the 'assembly' of this apparatus would lead to the bacteriophage resistant phenotype. More studies will be necessary to elucidate the nature of this interesting phenotype.

6.8.4. OutF homologues: functional roles

Studies on PulF have shown that this protein is essential for the secretion of pullulanase in K. oxytoca and is located in the IM (Pugsley et al., 1990a). The exact role of XpsF is not known although it is a part of the xps gene cluster and is thus implicated in the secretion of Pel and Cel from X. campestris (Dums et al., 1991). The gene encoding PilC (pilC) is downstream of pilB and is essential for the biogenesis of type IV pili in P. aeruginosa (Nunn et al., 1990). It was not possible to prove that any of the Ecc Out- mutants were allelic with outF. The relatedness of OutF to PulF strongly suggests that OutF is required for Pel and Cel secretion from Ecc. Computer predictions suggest that OutF might be strongly embedded in the IM (section 6.5.2.2.3.). However, localisation studies will be needed to confirm this.

6.8.5. OutG homologues: functional roles

OutG, PulG and XpsG all have N-terminal sequences which are a feature of immature type IV pilin monomers. The VirB3 protein from B. subtilis also displays some similarity with this group of proteins. This protein is essential for DNA binding by competent cells of this bacterium (Breitling and Dubnau, 1990). The N-terminal region of the pilin protein contains the MePhe cleavage site (Marrs et al., 1985). An N-methyl-phenylalanine residue is the first amino acid residue in the 'mature' pilin protein (Hermodson et al., 1978; Marrs et al., 1985). The MePhe cleavage site is usually preceded by a relatively short potential signal-sequence of only six or seven amino acids in a number of different Gram-negative bacteria including Moraxella bovis (Marrs et al., 1985), Moraxella nonliquefaciens (Froholm and Sletten, 1977), Bacteroides nodosus (Finney

et al., 1988; Mattick et al., 1991), Pseudomonas aeruginosa (Pasloske et al., 1988), Neisseria gonorrhoeae (Meyer et al., 1984) and Neisseria meningitidis (Potts and Saunder, 1988). The assembly of pilin monomers into pili by B. nodosus involves a further internal processing event in addition to the MePhe cleavage event (Elleman et al., 1986). This produces two non-covalently linked pilin subunits and is a result of a proteolytic cleavage event (Elleman et al., 1986). When expressed in E. coli, pilin monomers from P. aeruginosa were not polymerised into pili (Strom and Lory, 1986). This suggests that other factors (not present in E. coli) are required for pilus assembly (see section 6.8.3. and 6.8.4.).

The N-terminal regions of all the type IV pili monomers are all strongly conserved. This might be due to functional constraints involved in the assembly of pili from pilin monomers. Mutants of N. gonorrhoeae have been isolated which exhibited altered piliation (Koomey et al., 1991). A single amino acid residue exchange (Gly to Ser) at position -1 in the pre-pilin abolished piliation. In some cases the pilin subunits are interchangeable between different bacteria. For example type IV pilin monomers from M. bovis assemble into extracellular pili when their genes are expressed heterologously in P. aeruginosa (Beard et al., 1990). The predicted type IV pilin (TcpA) from Vibrio cholerae also has a predicted MePhe cleavage site. However, this site is preceded by 22 amino acids instead of the more typical six or seven amino acid residues (Faast et al., 1989). The putative MePhe cleavage site of the predicted OutG protein is preceded by 22 amino acid residues whereas the PulG protein has the usual six amino acid type IV pilin signal-sequence. This shows that OutG and PulG are similar in that they both share the MePhe cleavage consensus, but they differ with respect to the length of this signal-sequence. Further evidence that the OutG

protein is exported into the periplasm, or has periplasmic domains, comes from studies with the transposon TnphoA. A 'blue' Out- mutant of Ecc (PR33) generated using TnphoA was complemented by a small fragment of DNA containing the complete outG gene (D. Whitcombe, pers. comm.). This 'blue' fusion must have resulted from an in-frame insertion event of TnphoA into an exported or membrane located protein. Upon translation this event would lead to the production of a chimeric protein which directs the PhoA moiety to the periplasm where it is biologically active.

Pili are involved in bacterial processes such as motility and adhesion. Type IV pili have a polar location and are involved in a form of motility which is known as 'twitching motility' (Ottow, 1975; Henrichsen, 1983). Pili are retractile structures and in P. aeruginosa were found to be hollow (inner diameter of 12 Angstrom) and composed of an helical array of pilin subunits (Folkard et al., 1981). It is important to note that pili or pilin based structures might be involved directly in the uptake of DNA by N. gonorrhoeae and B. subtilis. N. gonorrhoeae is highly competent for its own DNA and is thought to use this process to catalyse recombination between copies of its pilin genes and thus enhance antigenic variation (Seifert et al., 1990). Indeed, mutations that abolish piliation have been shown to drastically reduce the transformation efficiency of such bacteria (Seifert et al., 1990). Work carried out on B. subtilis has shown that mutations in the ComG3 protein lead to a competent deficient phenotype (Breitling and Dubnau, 1990). This protein, along with ComG4 and ComG5, also shows N-terminal homology to the type IV pilins. However, pili have not been detected in this bacterium and it might appear unusual that a pilin monomer would be made by a non-piliated bacterium. These workers demonstrated that ComG3 was exposed on the outer surface of the cell

membrane and suggested that it might be involved in anchoring additional molecules to produce a structure (pili like ?) involved in competence.

6.9. A specialised eubacterial macromolecular trafficking apparatus

Some of the components of the macromolecular trafficking systems found in the bacteria discussed in this work exhibit remarkable similarities. Particularly noticeable is the strong conservation at both the levels of gene organisation and protein sequence between the protein secretion systems of Ecc (Out), K. oxytoca (Pul) and X. campestris (Xps). The similarities between these translocation systems suggests that they might have evolved from a common ancestor. It is interesting to speculate that the secretory systems of these different bacteria have evolved in order to translocate proteins unique to each bacterium.

The homologies between the Out, Pul and Xps proteins suggests that these secretion systems are similar but does not suggest anything about the possible functioning of these systems. The generation of more sequence data from the secretion systems of different bacteria might, however, help to identify regions of strong conservation which might result from evolutionary structural constraints.

The homologies between components of the protein translocation systems in Ecc, K. oxytoca and X. campestris with the DNA translocation systems of B. subtilis and A. tumefaciens and proteins of type IV pilin biogenesis were particularly important. It appears that at least one component (the OutE homologue) is common to all these different macromolecular transport systems. Strongly conserved in this predicted cytoplasmic protein (OutE) are potential nucleotide binding sites. It is interesting to speculate that this protein might somehow be necessary for

providing energy for the translocation process or perhaps the assembly of the translocation apparatus.

Macromolecules such as DNA or proteins might be able to pass through the centre of pili-like structures which somehow connect the IM and OM. Such a transport system might behave as a channel connecting the cytoplasm to the extracellular milieu. This might explain why pilin-like molecules are found in a number of different secretion apparatuses (Out, Pul, Xps and VirB) and import systems (ComG). Alternatively, the retractile properties of a pilus spanning the IM and OM might result in bringing these membranes into close proximity. Secreted proteins might traverse the IM (via the general export pathway) and immediately come into contact with the OM, resulting in their subsequent translocation to the extracellular milieu. This step would be expected to involve other discriminatory factors to prevent the loss of naturally periplasmic proteins. This hypothesis would still be consistent with the two-step hypothesis which is generally favoured for Out-like secretion. Furthermore, mutations in such a pathway might also be expected to accumulate periplasmic intermediates of the normally secreted proteins.

In summary, it has been shown that the Out proteins identified during the course of this work are members of a family of proteins which are involved in macromolecular trafficking in a diverse range of bacteria. This family of proteins probably represents the major generic trans-membrane targeting route out of Gram-negative prokaryotic cells, remaining undiscovered until now because it is absent in *E. coli*.

CHAPTER 7

IDENTIFICATION OF Out PROTEINS

7.1. Introduction

7.1.1. Gene expression and protein product identification

The identification of proteins encoded by cloned genes is an important technique in molecular biology. It is useful to determine the apparent molecular weight of proteins in order to correlate this with the predicted sizes of proteins determined by DNA sequence analysis. Some expression systems (e.g. *E. coli* maxicells or minicells) can also yield information on molecular processing events such as signal-sequence cleavage. Protein expression studies are also fundamentally important in enabling over-expression from cloned genes, either for industrial reasons or for production of large quantities of purified protein for raising antibodies or for X-ray crystallographic investigations.

Many techniques are available for identifying protein products encoded by cloned fragments of DNA. These methods include the use of *E. coli* maxicells (Sancar *et al.*, 1979), *E. coli* minicells (Clarke-Curtis and Curtis, 1983) and the cell-free coupled transcription/translation system (Zubay, 1973). It is sometimes necessary to control the expression of cloned genes. Some genes are weakly expressed for a variety of reasons. Their own promoters might be weak, they might be poorly expressed in a foreign host or they might have been cloned without their own promoter. Alternatively, the high level expression of some cloned genes might be toxic to the host cell. For these reasons it is often desirable to place cloned genes downstream of a powerful promoter which can be tightly regulated. Such promoters include the *trp-lac* (TAC) promoter (Amman *et al.*, 1983) or the bacteriophage T7 ϕ10 promoter (Tabor and Richardson, 1985). These expression systems allow gene expression in a controlled manner.

Once a gene has been expressed its product can be subjected to SDS polyacrylamide gel electrophoresis (SDS PAGE) and identified using numerous techniques. Gels from SDS PAGE procedures can be stained to reveal the total number of protein bands in a loaded sample. Such procedures utilise Coomassie blue or the Silver staining procedure. Specific protein bands may be detected by Western blotting if an appropriate antibody is available.

Some gene expression systems allow desired proteins to be specifically radio-labelled. Radio-labelled proteins are easily identified after SDS PAGE using autoradiography.

7.1.2. Visualisation of Out proteins

The aim of this section of work was to visualise the Out proteins using SDS PAGE and to determine their apparent molecular weights using known molecular weight standards. The apparent molecular weights of the Out proteins could then be compared with their predicted sizes as determined from the DNA sequence of *out* genes. Several gene expression systems were used in this study, and are described below. However, only one of these was successful. This was the T7 ϕ10 promoter-directed system (Tabor and Richardson, 1985) which will be described in detail.

7.2. In vitro coupled transcription/translation (Zubay) system

7.2.1. Introduction

The *in vitro* coupled transcription/translation system consists of a cell-free extract ('S' extract) and a cocktail of amino acids (minus methionine). Plasmid DNA (carrying the desired gene) is added to the above mix and translation is initiated by the addition S^{35} methionine (S^{35} Met).

Radioactively labelled methionine residues (^{35}S Met) are incorporated into any polypeptides produced under this regime. This approach was undertaken in order to identify the protein products encoded by out genes. The procedure followed was exactly as described by the manufacturers (Amersham International) (section 2.28.1.). Samples were examined using SDS PAGE (section 2.29.). Gels were treated with 'Amplify' (from Amersham International) according to the manufacturers instructions and dried (vacuum, 2 hr, 60°C, Biorad Gel drier) prior to autoradiography.

The cell-free nature of this system ensures that translocation dependent protein processing does not occur. This feature is useful for determining the size of immature forms of exported proteins.

7.2.2. Results and discussion

Using this system it was not possible to identify protein bands unique to any of the out+ carrying cosmids (cHIL122, cHIL220, cHIL159, cHIL208) (section 5.4.1.) or plasmid p3.7R1 (Figure 6.1.). Common bands were identified with mobilities consistent with vector encoded proteins B-lactamase (Bla) and the tetracycline resistance protein (Tet). It was possible that the out genes were not expressed by the E. coli gene expression machinery or were expressed at levels too low to detect in this system. The expression of out genes in Ecc could be due to the activity of a vector (pHC79) promoter. Such expression might be sufficient for complementation but not for protein detection in the Zubay system. Alternatively, the Out proteins might be unstable or rapidly degraded in this expression system. This might be due to the absence of cell membranes which might be required to stabilise membrane-localised Out proteins.

The results from this experiment are consistent with those by co-

workers in this laboratory when using E. coli maxicells to identify proteins encoded by the same out+ cosmids (S. Wharam and D. Whitcombe, pers. comm.). No unique protein bands were identified when samples from E. coli maxicell experiments were analysed by SDS PAGE followed by autoradiography.

Work is currently being undertaken in this laboratory in order to construct a recA- mutant of Ecc. Such a strain might be used as a natural Ecc maxicell host for studying the expression of out genes from cosmids/plasmids.

7.3. T7 gene 10 expression system

7.3.1. Introduction

It was not possible to identify out gene products using the in vitro coupled transcription/translation system (section 7.2.) or E. coli maxicells (as found by co-workers in this laboratory). Another approach taken was to study expression of the out genes when driven by the strong bacteriophage T7 010 promoter. An elegant system, designed by Tabor and Richardson (1985), allows tightly controlled, high level expression of cloned genes using the bacteriophage T7 010 promoter. The main features of the T7 system are the high specificity of T7 RNA polymerase for its own promoters and the insensitivity of T7 RNA polymerase (but not E. coli RNA polymerase) to the antibiotic rifampicin. These features allow exclusive expression of genes cloned downstream of the T7 010 promoter in the presence of T7 RNA polymerase and rifampicin. The principle behind the experimental protocol is summarised in Figure 7.1. The protein products of genes expressed from T7 promoters can be radio-labelled and identified using SDS PAGE followed by

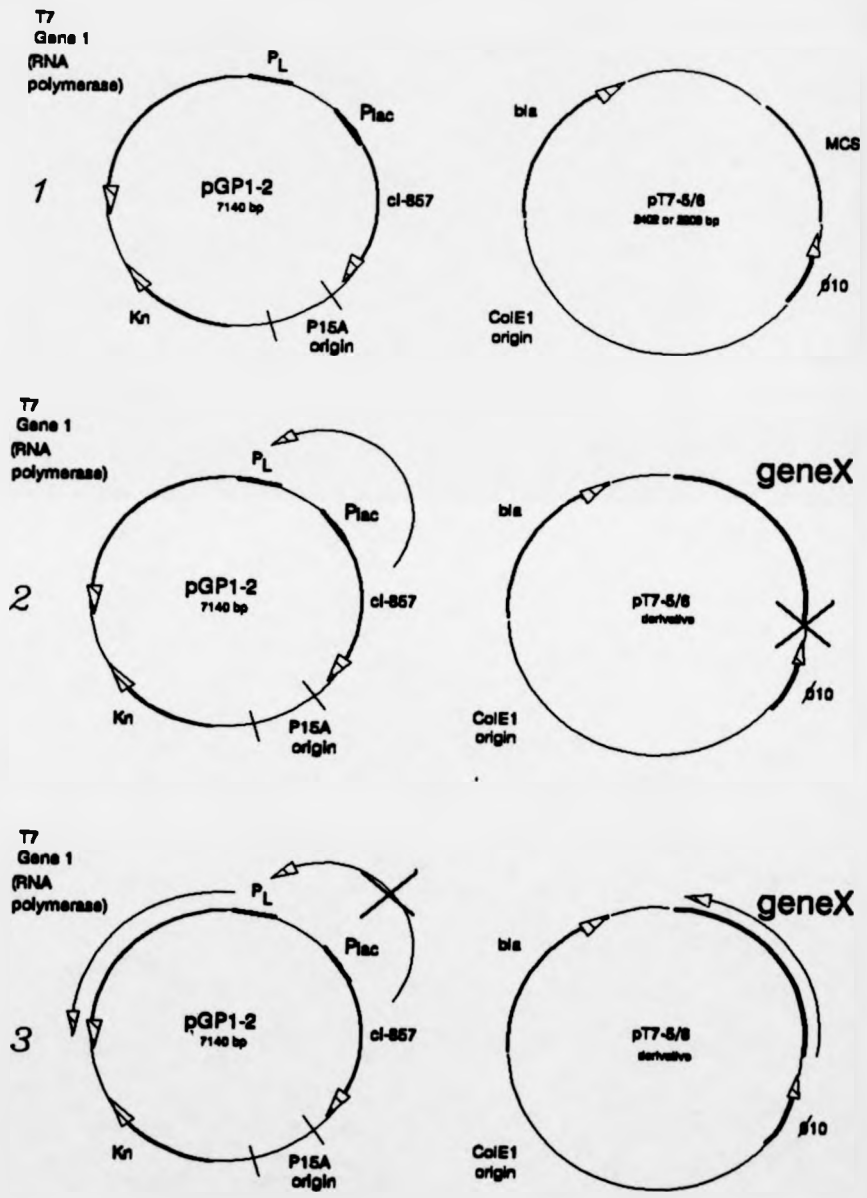
Legend to Figure 7.1.

1) The two plasmids used in the T7 expression system are pGP1-2 or pT7-5/6. Plasmids pT7-5 and pT7-6 are essentially the same except the multiple cloning site (MCS) is in the opposite orientation. Genes of interest are cloned into the MCS of either pT7-5 or pT7-6 so they are driven by the T7 ϕ10 promoter. Plasmids pGP1-2 and pT7-5/6 are compatible and can be introduced into the same E.coli host simultaneously.

2) Derivatives of pT7-5/6 carrying geneX are transformed into E.coli K38 (pGP1-2) and transformants are selected using Ap^r and Kn^r at 30°C. When transformants are grown at 30°C the cI-857 gene product binds to P_L and represses production of T7 RNA polymerase which is under control of P_L. Under these conditions geneX is not expressed.

3) The same cells as in 2) are shifted to 42°C. The cI-857 gene product is thermally denatured and can no longer bind to P_L. T7 RNA polymerase is produced which then recognises the T7 ϕ10 promoter on the pT7-5/6 derivative. This leads to the expression of geneX. E.coli RNA polymerase can be inhibited by the addition of rifampicin prior to labelling the geneX product with ^{35}S Met.

Figure 7.1. The T7 expression system



autoradiography. The strategy for this experiment is summarised below.

Plasmids pT7-5 and pT7-6 contain the T7 ϕ10 promoter downstream of a multiple cloning site (MCS). In each plasmid the orientation of the MCS is reversed. Restriction maps of these vectors are given in Appendix I. Derivatives of pT7-5/6 were introduced (by transformation) into E. coli K38 (pGP1-2) (section 2.26.). Plasmid pGP1-2 is compatible with pT7-5/6 and carries a heat inducible gene which encodes T7 RNA polymerase. Upon heat induction T7 RNA polymerase is produced which allows the specific expression of genes controlled by T7 promoters. Furthermore, the addition of rifampicin (after induction of T7 RNA polymerase) inhibits E. coli RNA polymerase. Protein products from genes expressed from the T7 ϕ10 promoter can now be labelled with S^{35} Met. Samples can be analysed by SDS PAGE followed by autoradiography. Experimental details of this method are outlined in section 2.28.2.2. and are summarised in Figure 7.1. The pT7-5/6 derivatives were made as described in section 2.28.2.1. and are described in Figure 7.2. Plasmid constructs were designed to contain regions of the out gene cluster which contained sequenced genes or lay within regions defining complementation groups. Recently, all these regions of DNA were sequenced in this laboratory. The genetic organisation of the out gene cluster is also shown in Figure 7.2. Proteins resulting from the T7 expression experiment were separated using SDS PAGE (section 2.29.). Gels were then subjected to autoradiography as described for the Zubay gel (7.2.).

7.3.2. Results

An autoradiograph showing Out proteins is presented in Figure 7.3. The plasmid constructs in lanes 1 to 4 were derivatives of either pT7-5 or pT7-6. In all cases the bacterial strain used was E. coli K38 containing

Figure 7.2 (opposite page)

Figure legend

Figure 7.2 shows the organisation of out genes in the out gene cluster and some of the restriction enzyme sites within this region. This entire DNA fragment was carried by cHIL208, a cosmid which partially restored the Out+ phenotype in all Out- mutants. The position of out genes and the direction of expression is indicated by an arrow. The capital letter beneath each arrow denotes the out gene. Four pT7-5/6 (which contain the T7 ϕ10 promoter) derivatives were made which contained regions of the out gene cluster. These are described below. Truncated genes are indicated by ' or *.

p308/1
HindIII- BamHI fragment from cosmid cHIL208 as indicated in Figure 7.2. This DNA fragment was cloned into pT7-6 which had been digested with SalI and BamHI. This construct contains outD, outE, outF and a truncated outG (outG').

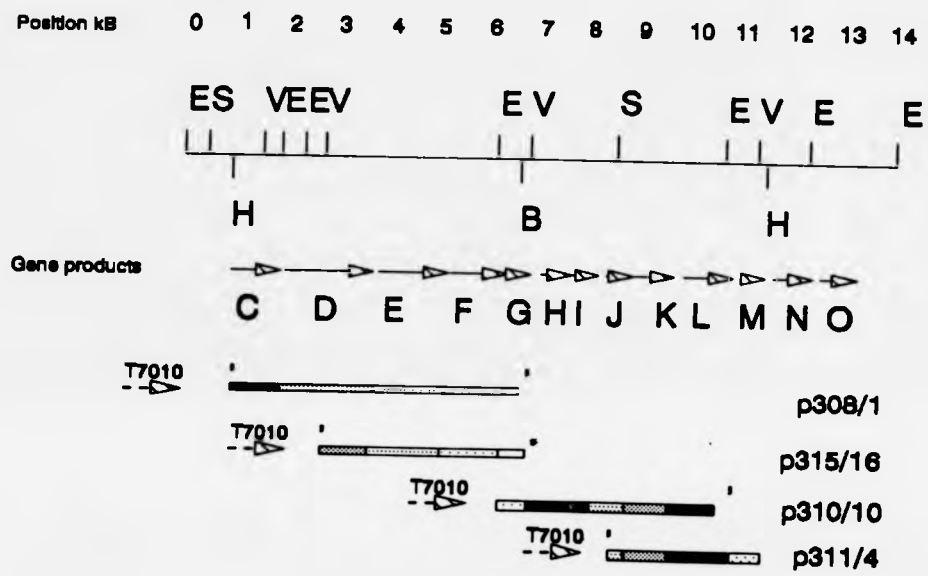
p315/16
EcoRV fragment from cosmid cHIL122 (section 5.4.), as indicated in Figure 7.2, cloned into the SmaI site of pT7-5. Constructs with the out fragment in the correct orientation were identified by restriction mapping. Plasmid p315/16 contains a 5' truncate of outD ('outD), outE, outF and a truncated outG (outG*).

p310/10
Contains the entire EcoRI fragment of cHIL220 (section 5.4.) as indicated in Figure 7.2. This EcoRI fragment was cloned into the EcoRI fragment of pT7-5. Plasmids with the insert in the correct orientation were identified by restriction mapping using EcoRV which cuts the DNA insert assymetrically. This construct contains outG, outH, outI, outJ, outK and a truncated outL (outL').

p311/4
SalI - HinDIII fragment from plasmid pHIL159/3 (Figure 5.8.), as indicated in Figure 7.2 cloned into pT7-5 which had been cut with the same restriction enzymes. This construct contains a truncated outJ ('outJ), outK, outL and outM. The translation termination point of outM is at the HindIII site.

The cloning vectors (pT7-5/6) are given in appendix I.

Figure 7.2 Construction of clones for T7 gene 10 promoter expression studies



Key for Figure 7.3.

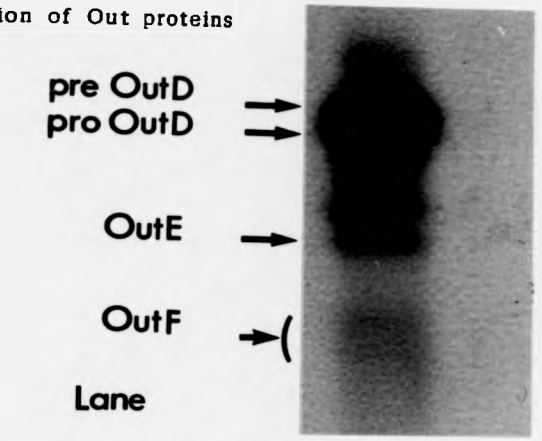
Lane M	Molecular weight standards (sizes in kD)
Lane 1	p311/4
Lane 2	p310/10
Lane 3	p315/16
Lane 4	p308/1
Lane 5	pT7-5 (uninduced)
Lane 6	pT7-5 (induced + rifampicin)
Lane 7	pT7-5 (induced - rifampicin)

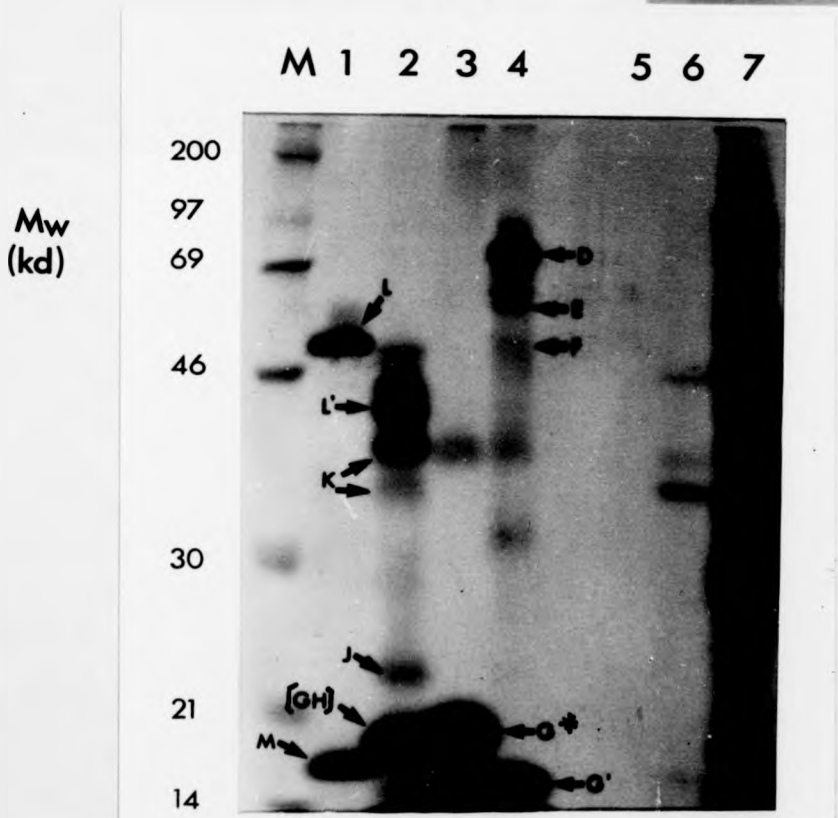
Figure 7.3. legend

The plasmid constructs and their potential gene products are shown in Figure 7.2. The sizes of the molecular weight standards (in kD) are printed on the left of the gel photograph. The proteins bands most likely to correspond to predicted Out proteins are indicated with an arrow and a letter for the protein name. Truncated proteins resulting from cloning are indicated with apostrophes. Some truncated Out proteins contained vector encoded amino acids as a result of readthrough into the vector. Other bands could not be assigned to any Out proteins. The sizes of Out proteins, as calculated from this gel, are given in Table 7.1. and compared with possible corresponding Out proteins predicted from the DNA sequence

The inset shows the upper region of lane 4 magnified. This shows more clearly that OutD could be a doublet and may exist as pre and pro forms.

Figure 7.3 Identification of Out proteins





pGP1-2 and pT7-5/6 derivatives. The known molecular weight standards ('Rainbow markers' purchased from Amersham International) were loaded in the track labelled M. Experimental controls were loaded in lanes 5, 6 and 7. Track 5 demonstrated that radio-labelled proteins were not produced when the system was uninduced, that is when there was no temperature shift to 42°C. Track 6 shows the same strain as used for track 5 but in this case induced using a temperature shift. Some radio-labelled proteins can be seen in this sample which appear to correspond to highly expressed cellular background proteins as seen in track 7. Track 7 shows the level of cellular protein production after temperature induction in the absence of rifampicin. The addition of rifampicin drastically lowers the levels of cellular background proteins. This is clearly demonstrated by comparing lanes 6 and 7.

Lanes 1 to 4 contain protein bands which are absent in the induced vector (without insert) control (track 6). These radio-labelled proteins must be encoded by the fragments of DNA from the out gene cluster. The sizes of the expressed proteins were calculated from their migration distances using a molecular weight standard calibration curve.

Lanes 1 and 2 contained proteins encoded by out genes which had not been sequenced at the time of this experiment. This region of DNA had been used to complement Out- mutants. The Out proteins encoded by these regions overlapping regions of DNA were clearly visible. In lane 1 (p311/4) proteins calculated to be 50.0 kD and 17.0 kD were present. Lane 2 (p310/10) contained unique bands with relative mobilities corresponding to molecular weights of [49.0 kD], 42.0 kD, [45.0 kD], 38.0 kD and 23.0 kD. A further diffuse but intense band was present which migrated to a position consistent with a 18.5 kD protein.

Lanes 3 and 4 revealed the presence of proteins expressed from

overlapping and sequenced regions of the out gene cluster. Lane 3 (p315/16) had one diffuse but very intense band migrating at 20.0 kD. Lane 4 (p308/1) contained proteins corresponding to molecular weights of 75.0 kD, 70.0 kD, [60.0 kD], 56.0 kD, 51.0 kD, [31.5 kD] and 15.0 kD. The protein bands which were enclosed by square brackets [] could not be assigned to predicted proteins. These could be degradation products of larger proteins or might be due to alternative translational starts. These results are summarised in Table 7.1. This table also shows the predicted sizes of out gene products determined from DNA sequence analysis of the out gene cluster.

7.3.3. Discussion

Using the T7 promoter expression system it was possible to identify proteins encoded by out genes.

Plasmids p315/16 and p308/1 were designed to contain DNA from the sequenced region of the out gene cluster. The predicted sizes of proteins encoded by p308/1 and p315/16 were compared with the actual sizes of proteins as visualised in this experiment and are shown in Table 7.1. In this table the 'observed' proteins most likely to correspond to the sizes of the predicted proteins are given. In order to confirm that a predicted protein corresponded to a particular protein band on the gel it would have been necessary to clone each individual gene into the expression plasmid. This, however, was not possible due to the lack of appropriate restriction sites within the out gene region.

The largest Out protein, OutD, was identified as an intense doublet. The predicted OutD protein contained a classical E. coli signal-sequence (Figure 6.3.). OutD appears to be recognised by the E. coli export machinery. The apparent molecular weight of OutD is 75 kD (immature) and

Table 7.1. Predicted and apparent sizes of Out proteins

Out protein	Predicted size (kD) (from DNA sequence)	Observed size (kD) (from Fig.7.3)	Plasmid construct
pre-OutD	71.279	75	p308/1
pro-OutD	68.720	70	p308/1
OutE	55.281	56	p308/1
OutF	45.167	51	p308/1
OutG'	12.615 (13.997)	15	p308/1
OutG*	14.751 (16.804)	20	p315/16
OutG	17.353	18[1]	p310/10
OutH	21.392	18[1]	p310/10
OutI	13.841	-[3]	p310/10
OutJ	24.853	23	p310/10
OutK	36.573	38[2]	p310/10
		45[2]	p310/10
OutL'	42.275 (44.580)	42	p310/10
OutL	47.644	50	p311/4
OutM	18.722	17	p311/4

Truncated gene products are illustrated with ' or *. The predicted and apparent sizes of truncated proteins are shown. The translation of truncated Out proteins is terminated by vector stop codons. The total size of Out/Vector hybrid proteins is given in brackets.

1) Out proteins G and H were encoded by the same plasmid construct (p310/10). OutG and OutH might not separate on this gel because of the intensity of the OutG protein band (see lanes 3 and 4) and the similarity in the predicted sizes of these two proteins.

2) Two proteins which might correspond to OutK are shown.

3) OutI, with a predicted size of 13.8 kD probably ran off the end of the gel.

70 kD (mature). A band migrating at 60 kD is also presented which could represent a product of OutD after a further specific processing event or might be simply a breakdown product of OutD. The corresponding secretion protein from K. oxytoca, PulD, was identified by immunoblotting fractionated cells with a PulD-specific antibody (d'Enfert et al., 1989). PulD was located predominantly in the OM but was also found in the IM, especially in cells carrying pulD on a multicopy plasmid. The size of PulD in E. coli minicells was 68 kD which is in close agreement with the predicted size of 67.666 kD. Workers investigating an Ech 3937 Out- mutant (OutJ) identified a complementing gene named outJ. The protein encoded by outJ had a molecular weight of 83 kD (immature) and 81 kD (mature) (Ji, et al., 1989). Furthermore, in the OutJ mutant, a 68 kD periplasmic protein was absent. Ji et al. (1989) suggested that OutJ might be processed in several steps from 83 kD to 81 kD and finally 68 kD (corresponding to the protein lacking in the OutJ mutant). The N-terminal region of outJ has been sequenced (J. Ji, pers. comm.) and from this limited information the predicted protein appears similar to OutD from Ecc and PulD from K. oxytoca (d'Enfert et al., 1989). Therefore it seems likely that OutD (Ecc), OutJ (Ech) and PulD (K. oxytoca) are homologous.

A protein band with an apparent size of 56 kD was identified. This might correspond to OutE which has a predicted size of 55.281 kD. The corresponding K. oxytoca secretion protein, PulE, has been identified using E. coli minicells and a T7 expression system, and was located in the cytoplasm (Pugsley et al., 1990a).

A faint protein band migrating at 51 kD was identified. This faint band was diffuse, which might suggest that it was produced at low levels or/and degraded. This protein band might be OutF. OutF is predicted to be

located in the IM of Ecc (section 6.5.2.2.3.) and has a predicted molecular weight of 45.167 kD. If OutF is located in the IM it might be difficult to remove from the IM prior to SDS PAGE.

An intense but diffuse protein band with a mobility of 15 kD was identified corresponding to a truncated version of OutG (OutG') which has a predicted size of 12.165 kD. The intensity of this band suggests that outG is highly expressed compared to other Out proteins. OutG protein shares structural similarities to pilin monomers from a number of different bacterial species (section 6.7.2.4.). It could possibly be a building block (monomer) of a structure (pilus like?) involved in secretion. A full size version of the OutG protein is evident in lane 2 of Figure 7.3. This intense but diffuse band had a mobility of 18 kD which is in good agreement with the size predicted from the DNA sequence (17.353 kD).

Lane 3, plasmid construct p315/16 had only one unique protein band migrating at 20 kD. This was probably the truncated/hybrid version of OutG* which includes some vector-encoded amino acid residues. Interestingly, plasmid p315/16 (lane 3) should also encode the complete OutE (55.281 kD) and OutF (45.167 kD) as well as OutG*. However, proteins corresponding to OutE and OutF were absent. This result indicates that the expression of outE and outF might be affected when outD is not expressed. This might be due to a regulation mechanism at the level of translation. As mentioned in section 6.4.3., the reading frames of outD, outE and outF overlap indicating that their expression might be translationally coupled. Translational coupling occurs when reading frames overlap ,or are separated by five nucleotides or less (Gold and Stormo, 1987). In a translationally coupled system the initiation of translation of the downstream protein is thought to be achieved without the ribosome detaching from the mRNA. This negates the need for a

Shine Dalgano sequence in downstream translationally coupled genes. Translational coupling is a mechanism which is common in large operons and might be involved in stabilising the mRNA molecule (Gold and Stormo, 1987). This mechanism might also be involved in regulating the levels of out gene expression and may ensure that Out proteins are synthesised in stoichiometrically appropriate amounts.

The entire out gene cluster has recently been sequenced in this laboratory. This includes the out DNA in the plasmid constructs p308/1 and p310/10. The sequence of outH - outO will not be discussed in this work, However, the predicted sizes of some of these out genes is presented so they can be compared to Out proteins visualised in this work.

Plasmid p310/10 is predicted to encode OutG, OutH, OutI, OutJ, OutK and a truncated OutL (OutL'). A protein band corresponding to OutG and/or OutH was identified at 18 kD. Protein bands migrating at 23 kD (OutJ), 38,45 kD (OutK) and 42 kD (OutL') were identified. OutI (13.814 kD) was not visualised but as the lowest molecular weight marker was 14 kD, the OutI protein probably ran off the gel.

A protein of 50 kD encoded by p311/4 was probably the complete OutL (expected size of 48.864 kD) and another band of molecular weight 17.0 kD was probably OutM (expected size of 17.459 kD). OutN and OutO have not been cloned into the T7 expression vectors.

Further studies will be necessary to confirm the identity of Out proteins identified in this study. This preliminary investigation has demonstrated that the T7 expression system will be a useful tool for expressing out genes and investigating the Out proteins. It will be necessary to clone the out genes onto smaller DNA fragments in order to accurately assign Out proteins to particular out genes. However, this approach might be

hampered by the translational coupling of some of the out genes and the unavailability of restriction sites. Gene interruption experiments would also help in assigning Out proteins to out genes. This would be done by inserting transposons/interposons into known out genes and monitoring the effect on the out gene product. Another approach would be to make a series of deletions in the out encoding DNA (from the 3' end) and monitoring the effects that certain deletions had on the Out proteins produced.

It would be interesting to identify the exact cellular location of the Out proteins when expressed in E. coli. This system might also be useful in the over-production of Out proteins. The large quantities of purified proteins produced would enable Out-specific antibodies to be raised which could then be used in localisation studies.

Recently the T7 expression system has been reconstructed in Ecc (P. Douglas, pers. comm.). This has been used to identify the Out products encoded by out genes carried by p308/1 (as used in this study). Work is now being undertaken to determine the cellular locations of Out proteins in Ecc using the T7 expression system in conjunction with cell fractionation techniques.

In summary, the T7 expression system has been used successfully for the expression of out genes allowing a tentative identification of out gene products. Preliminary studies have allowed the identification of OutD, OutE, OutF and OutG and supported the predictions that translational coupling might be a factor in the expression of some of the out genes. Future work will concentrate on identifying the remaining Out proteins and determining their cellular locations in both Ecc and E. coli. Long term aims will be to over-express the out genes and purify the Out proteins. It will then be possible to raise antibodies against the Out proteins. Anti Out antibodies

would be useful for determining the cellular locations of Out proteins and might help to identify the assembled secretion apparatus in Ecc.

CHAPTER 8

GENERAL DISCUSSION

8.1. Introduction

At the start of this research project very little was known about protein secretion by Gram-negative bacteria. Virtually nothing was known about the secretion of extracellular enzymes by Erwinia, although mutants defective in extracellular enzyme secretion (Out-) had been isolated (Andro et al., 1984). Work by Forbes and Perombelon (1985) and Hinton (1986) led to the development of a genetic system for studying Ecc. These workers had developed the basic genetic tools which were essential in carrying out the work described in this thesis. Using these genetic tools, the aim was to effect a molecular genetic analysis of enzyme secretion in Ecc. A summary of the major findings of this work is given below.

8.2. Summary of major findings from this work

The major findings from this work are as follows:

1) Mutants of Ecc were isolated which were defective for the production of extracellular enzymes. Some classes of mutant were pleiotropically defective in the production of two or more extracellular enzymes (chapter 3).

2) One class [Pel-, Cel-, Prt+] was shown to accumulate Pel and Cel intracellularly but produce Prt as normal. Mutants of this class have been called Out- mutants (chapter 4).

3) The Out- class of mutants accumulated Pel and Cel within the periplasm (chapter 4).

4) One Out- mutant (RJP190) was partially resistant to infection by two Ecc bacteriophages, ØD-2 and Ø565 (chapter4).

5) An Ecc cosmid library was used to complement 12 of the 14 EMS Out- mutants and restore the Out+ phenotype (chapter 5).

6) The out+ cosmids were categorised into three groups (a, b and c) on the basis of their complementation pattern and using restriction endonuclease mapping (chapter 5).

7) Experiments were carried out to subdivide the group C mutants into two groups. Collaboration with another worker (D. Whitcombe) led to the identification of six complementation groups, and the production of a restriction map for the 12 kb region of DNA containing out+ genes (chapter 5).

8) The nucleotide sequence of a 3.8 kb region of the out gene cluster was determined. This fragment overlapped at both the 5' and 3' ends with DNA fragments sequenced by co-workers in this laboratory. When analysed, the contiguous 5.7 kb fragment of DNA (including the 3.8 kb region) contained four orfs (coding for OutD, OutE, OutF and OutG) (chapter 6).

9) Several out gene products (including OutD, OutE, OutF and OutG) were identified. This was achieved by expressing out genes under the control of a T7 Ø10 promoter (chapter 7).

10) The putative secondary structures and other possible features of the Out proteins were analysed. OutD was predicted to be synthesised as a precursor with an N-terminal signal-sequence and be located in the OM. OutE was predicted to be cytoplasmic and have potential nucleotide binding domains. OutF was predicted to contain strong membrane anchors and be located in the IM. OutG was predicted to contain an N-terminal type IV pilin-like signal-sequence (chapter 6).

11) Similarities were found between the Out proteins and proteins performing macromolecular trafficking functions in a wide range of bacteria, including Pul (secretion of pullulanase from K. oxytoca), Xcp (secretion of numerous extracellular enzymes from P. aeruginosa), Xps (secretion of pectinases, cellulases and amylases from X. campestris), VirB (involved in Ti DNA transfer out of A. tumefaciens), ComG (DNA uptake by B. subtilis) and pIV (filamentous bacteriophage protein involved in phage release) (chapter 6).

8.3. Concluding remarks and future work

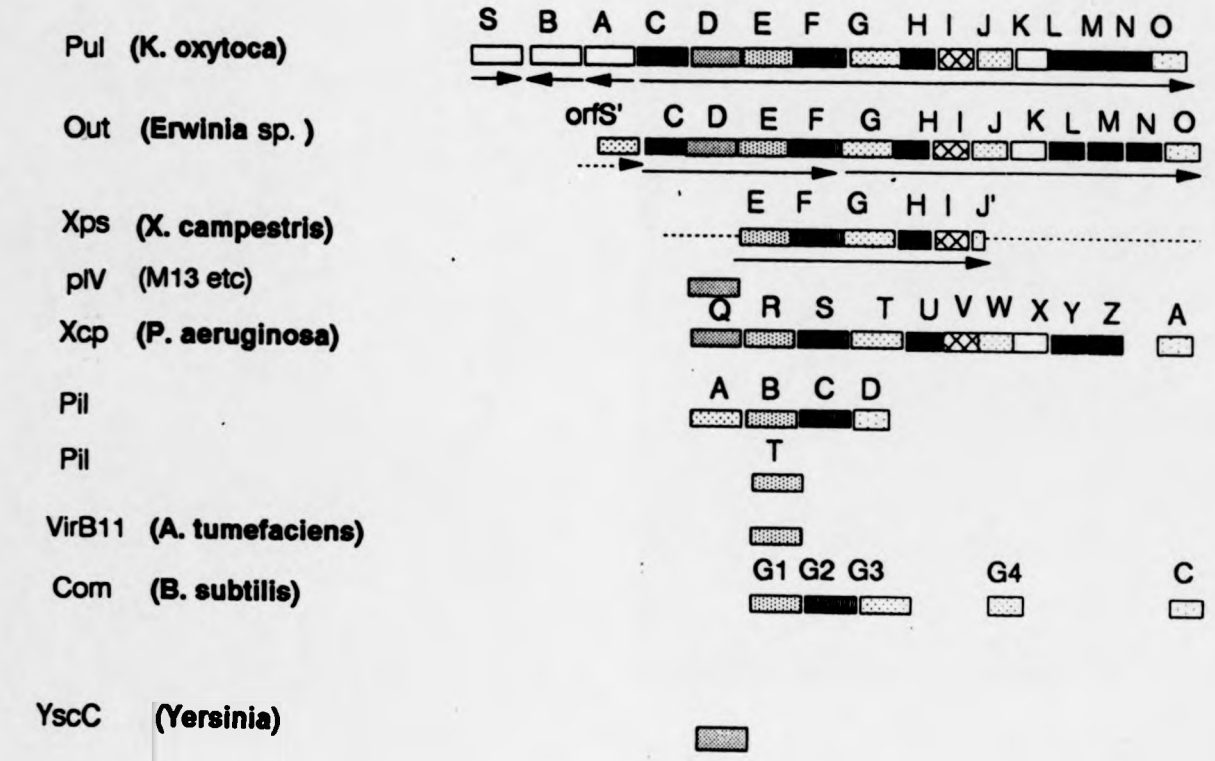
During the course of this work our understanding of the Ecc secretion apparatus has increased enormously. We now have a detailed understanding of some of the components of the Ecc secretion machinery. This demonstrates how powerful molecular genetic approaches can be for answering biological questions. Although the initial aim of this work was devoted entirely to investigating Erwinia, it is becoming apparent that proteins similar to the Out proteins discovered in this bacterium are widespread in Gram-negative bacteria, and are used for macromolecular trafficking both into and out of the bacterial cell. An up to date diagram showing the Out-like family of proteins and their sources is given in Figure 8.1. It is likely that this family

of proteins represents the components of the major pathway used for macromolecular trafficking by eubacteria.

Future work in this area will be extremely challenging. The components of the Out-like machinery have been defined in various systems (see Figure 8.1.); attention must now be directed to investigating the functioning of these components. Questions which need to be addressed are as follows: Do the Out-like proteins interact with each other, and if so what is the nature of this interaction? Do other Out-like components assemble into a structural secretion apparatus? What is the nature of the interaction between the secretion apparatus and the secreted proteins? Will the secretion apparatus of one genus recognise and secrete heterologous proteins? The limited knowledge we have suggests that the Out-like systems from different bacteria are specific for their own proteins (Py *et al.*, 1991). Experiments combining different components of the secretion apparatus from different bacteria might help us to gain an insight into the functional interaction between these components.

To conclude and reiterate, the Out-like family of proteins probably represents the major generic trans-membrane translocation apparatus of Gram-negative bacteria. The fact that this is probably the major route out of the Gram-negative cell begs the question why it has not been discovered until recently. Almost certainly this is because *E. coli* does not appear to have this pathway!

Figure 8.1. Membrane traffic proteins



Legend for Figure 8.1

This diagram was adapted from information supplied by A. Pugsley (pers. comm.)

The members of this family of membrane traffick proteins have been described in chapter 6 except for the following. The K. oxytoca Pul proteins PulS, PulB, PulA, PulH, PulI, PulJ, PulK, PulI, PulM, PulN and PulO have been described in review by Pugsley (1990). The genes encoding predicted Out proteins from Ecc have been sequenced in this laboratory. The genes encoding the Ech Out proteins OutH, OutI, OutJ and OutK have recently been sequenced (He et al., 1991a). The genes encoding the predicted Xcp proteins from P. aeruginosa have been sequenced in the laboratory of A. Lazdunski (pers. comm.). The XcpA (encoded by xcpA formerly xpx-1) has recently been published (Bally et al., 1991). The YscG protein is encoded as part of the virC operon from Yersinia enterocolytica which is involved in the secretion of Yop proteins (Michiels et al., 1991). ComC is involved in the genetic competence of B. subtilis (Mohan et al., 1989)..

REFERENCES

Albano,M., Breitling,R. and Dubnau,D.A. (1989) Nucleotide Sequence and Genetic Organization of the Bacillus subtilis comG Operon. J. Bacteriol. 171: 5386-5404.

Alcorn,S.M., Orum,T.V., Steigerwalt,A.G., Foster,J.L.M., Fogleman,J.C.and Brenner,D.J. (1991) Taxonomy and Pathogenicity of Erwinia cacticida sp. nov. Int. J. Syst. Bacteriol. 41: 197-212.

Allen,C., Stromberg,V.K., Smith,F.D., Lacy,G.H. and Mount,M.S. (1986) Complementation of an Erwinia carotovora Subsp. carotovora Protease Mutant with a Protease-Encoding Cosmid. Mol. Gen. Genet. 202: 276-279.

Altman,E., Kumamoto,C.A. and Emr,S.D. (1991) Heat-Shock Proteins can Substitute for SecB Function During Protein Export in Escherichia coli. EMBO J. 10: 239-245.

Amann,E., Brosius,J. and Ptashne,M. (1983) Vectors Bearing a Hybrid trp-lac Promoter Useful for Regulated Expression of Cloned Genes in Escherichia coli. Gene 25: 167-178.

Andro,T., Chambost,J., Kotoujansky,A., Cattaneo,J., Bertheau,Y., Barras,F., Van Gijsegem,F. and Coleno,A. (1984) Mutants of Erwinia chrysanthemi Defective in Secretion of Pectinase and Cellulase. J. Bacteriol. 160: 1199-1203.

Arai,S. and Watanabe,M. (1986) Freeze Texturing of Food Materials by Ice-Nucleation with the Bacterium Erwinia ananas. Agric. Biol. Chem. 50: 169-175.

Arbige,M.V. and Pitcher,W.H. (1989) Industrial Enzymology: A Look Towards the Future. TIBTECH 7: 330-335.

Bachmann,B.J. and Low,K.B. (1980) Linkage Map of Escherichia coli K12. Microbiol. Rev. 44: 1.

Bally,M., Ball,G., Badere,A. and Lazdunski,A. (1991) Protein Secretion in Pseudomonas aeruginosa: The xcpA Gene Encodes an Integral Inner Membrane Protein Homologous to Klebsiella pneumoniae Secretion Function Protein PulO. J. Bacteriol. 173: 479-486.

Bally,M., Wretlind,B. and Lazdunski,A. (1989) Protein Secretion in Pseudomonas aeruginosa: Molecular Cloning and Characterization of the xcp-1 Gene. J. Bacteriol. 171: 4342-4348.

Bankier,A.T., Weston,K.M. and Barrell,B.G. (1986) Random Cloning and Sequencing by the M13/Dideoxynucleotide Chain Termination Method. M.R.C. Laboratory of Molecular Biology, Cambridge.

Barny,M.A., Guinebretiere,M.H., Marcais,B., Coissac,E., Paulin,J.P. and Laurent,J. (1990) Cloning of a Large Gene Cluster Involved in Erwinia amylovora CFBP1430 Virulence. Mol. Microbiol. 4: 777-786.

Barras,F., Chambost,J.P. and Chippaux,M. (1984) Cellobiose Metabolism in Erwinia: Genetic Study. Mol. Gen. Genet. 197: 486-490.

Barras,F., Thurn,K.K. and Chatterjee,A.K. (1986) Export of Erwinia chrysanthemi (EC16) Protease by Escherichia coli. FEMS Microbiol. Lett. 34: 343-348.

Bassford,P., Beckwith,J., Ito,K., Kumamoto,C., Mizushima,S., Oliver,D., Randall,L., Silhavy,T., Tai,P.C. and Wickner,W. (1991) The Primary Pathway of Protein Export in E. coli. Cell 65: 367-368.

Bateman,D.F. and Basham,H.G. (1986) Degradation of Plant Cell Walls and Membranes by Microbial Enzymes. In: Encyclopedia of Plant Physiology, Vol. 4. Heitefuss,R. and Williams,P.H. (eds), Springer-Verlag: New York, pp 316-355.

Beard,M.K.M., Mattick,J.S., Moore,L.J., Mott,M.R., Marrs,C.F. and Egerton,J.R. (1990) Morphogenetic Expression of Moraxella bovis Fimbriae (Pili) in Pseudomonas aeruginosa. J. Bacteriol. 172: 2601-2670.

Beaulieu,C. and van Gijsegem,F. (1990) Identification of Plant-Inducible Genes in Erwinia chrysanthemi 3937. J. Bacteriol. 172: 1569-1575.

Beck,E. and Zink,B. (1981) Nucleotide Sequence and Genome Organisation of Filamentous Bacteriophages f1 and fd. Gene 16: 35-58.

Beck,E., Sommer,R., Auerswald,E.A., Kurz,C., Zink,B., Osterburg,G., Schaller,H., Sugimoto,K., Sugisaki,H., Okamoto,T. and Takanami,M. (1978) Nucleotide Sequence of Bacteriophage fd DNA. Nucl. Acids Res. 5:4495-4505.

Bieker,K.L. and Silhavy,T.J. (1990) The Genetics of Protein Secretion in E. coli. TIG 6: 329-334.

Biely,P., Mislovicova,D. and Toman,R. (1985) Soluble Chromogenic Substrates for the Assay of Endo-1,4-Beta-Xylanases and Endo-1,4-Beta-Glucanases. Anal. Biochem. 144: p 147-151.

Blight,M.A. and Holland,I.B. (1990) Structure and Function of Haemolysin B, P-glycoprotein and Other Members of a Novel Family of Membrane Translocators. Mol. Microbiol. 4: 873-880.

Blobel,G. and Dobberstein,B. (1975) Transfer of Proteins Across Membranes. II. Reconstitution of Functional Rough Microsomes from Heterologous Components. J. Cell Biol. 67: 852-862.

Bolivar,F. and Backman,K. (1979) Plasmids of Escherichia coli as Cloning Vectors. Meth. Enzymol. 68: 245-267. .

Borchert,T.V. and Nagarajan,V. (1991) Effect of Signal Sequence Alterations on Export of Levansucrase in Bacillus subtilis. J. Bacteriol. 173: 276-282.

Boyer,M.H., Cami,B., Chambost,J.-P., Magnan,M. and Cattaneo,J. (1987) Characterization of a New Endoglucanase from Erwinia chrysanthemi. Eur. J. Biochem. 162: 311-316.

Boyer,M.H., Chambost,J.P., Magnan,M. and Cattaneo,J. (1984) Carboxymethyl-Cellulase from Erwinia chrysanthemi. I. Production and Regulation of Extracellular Carboxymethyl-Cellulase. J. Biotechnol. 1: 229-239.

Boyer,M.H., Chambost,J.P., Magnan,M. and Cattaneo,J. (1984) Carboxymethyl-Cellulase from Erwinia chrysanthemi. II. Purification and Partial Characterization of an Endo-β-1,4-Glucanase. J. Biotechnol. 1: 241-252.

Bradley,D.E. (1980) A Function of Pseudomonas aeruginosa PAO Pili: Twitching Motility. Can. J. Microbiol. 26: 146-154.

Breitling,R. and Dubnau,D. (1990) A Membrane Protein with Similarity to N-Methylphenylalanine Pilins is Essential for DNA Binding by Competent Bacillus subtilis. J. Bacteriol. 172: 1499-1508.

Brissette,J.L. and Russel,M. (1990) Secretion and Membrane Integration of a Filamentous Phage-Encoded Morphogenetic Protein. J. Mol. Biol. 221: 565-580.

Brissette,J.L., Russel,M., Weiner,L. and Model,P. (1990) Phage Shock Protein, a Stress Protein of Escherichia coli. Proc. Natl. Acad. Sci. USA 87: 862-866.

Brissette,J.L., Weiner,L., Ripmaster,T.L. and Model,P. (1991) Characterization and Sequence of the Escherichia coli Stress-Induced psp Operon. J. Mol. Biol. 220: 35-48.

Brundage,L., Hendrick,J.P., Schiebel,E., Driessen,A.J.M. and Wickner,W. (1990) The Purified E. coli Integral Membrane Protein SecY/E is Sufficient for Reconstitution of SecA-Dependent Precursor Protein Translocation. Cell 62: 649-657.

Chambost,J.P., Boyer,M.H., Cami,B., Barras,E. and Cattaneo,J. (1985) Erwinia Cellulases. In Proc. 6th Int. Conf. Plant Path. Bact. Civerolo,E.L., Collmer,A., Gillaspie,A.G. and Davis,R.E. (eds). Dordrecht: Martinus Nijhoff, pp 150-159.

Charbit,A., Gehring,K., Nikaido,H., Ferenci,T. and Hofnung,M. (1988) Maltose Transport and Starch Binding in Phage-Resistant Point Mutants of Maltoporin. J. Mol. Biol. 201: 487-496.

Chatterjee,A., McEvoy,J.L., Chambost,J.P., Blasco,F. and Chatterjee,A.K. (1991) Nucleotide Sequence and Molecular Characterization of pnlA, the Structural Gene for Damage-Inducible Pectin Lyase of Erwinia carotovora Subsp. carotovora 71. J. Bacteriol. 173: 1765-1769.

Chatterjee,A.K., Ross,L.M., McEvoy,J.L. and Thurn,K.K. (1985a) pULB113, an RP4::mini-Mu Plasmid, Mediates Chromosomal Mobilization and R-Prime Formation in Erwinia amylovora, Erwinia chrysanthemi, and Subspecies of Erwinia carotovora. Appl. Environ. Microbiol. 50: 1-9.

Chatterjee,A.K., Thurn,K.K. and Tyrrell,D.J. (1981) Regulation of Pectolytic Enzymes in Soft Rot Erwinia. In Proc. 5th Int. Conf. Plant Path. Bact. Lozano,J.C. (ed). Cali, Columbia: CIAT, pp252-262.

Chatterjee,A.K., Thurn,K.K. and Tyrell,D.J. (1985b) Isolation and Characterization of Tn5 Insertion Mutants of Erwinia chrysanthemi that are Deficient in Polygalacturonate Catabolic Enzymes Oligogalacturonate Lyase and 3-Deoxy-D-Glycero-2,5-Hexodiulosonate Dehydrogenase. J. Bacteriol. 162: 708-714.

Chou,P.Y. and Fasman,G.D. (1978) Prediction of the Secondary Structure of Proteins from their Amino Acid Sequence. Advan. Enzymol. 47: 45-148.

Christie,P.J., Ward,J.E.,Jr., Gordon,M.P. and Nester,E.W. (1989) A Gene Required for Transfer of T-DNA to Plants Encodes an ATPase with Autophosphorylating Activity. Proc. Natl. Acad. Sci. USA 86: 9677-9681.

Clarke-Curtis,J.E. and Curtis,R. (1983) Analysis of Recombinant DNA Using E. coli mini-cells. Meth. Enzymol. 101: 342-362.

Clement,J.M., Perrin,D. and Hedgpeth,J. (1982) Analysis of λ Receptor and β-lactamase Synthesis and Export Using Cloned Genes in a Minicell System. Mol. Gen. Genet. 185: 302-310.

de Cock,H. and Tommassen,J. (1991) Conservation of Components of the Escherichia coli Export Machinery in Prokaryotes. FEMS Microbiol. Lett. 80: 195-200.

Collier,D.N., Bankaitis,V.A., Weiss,J.B. and Bassford,Jr.,P.J. (1988) The Antifolding Activity of SecB Promotes the Export of the E. coli Maltose-Binding Protein. Cell 53: 273-283.

Collmer,A. and Bateman,D.F. (1981) Impaired Induction and Self-Catabolite Repression of Extracellular Pectate Lyase in Erwinia chrysanthemi Mutants Deficient in Oligogalacturonide Lyase. Proc. Natl. Acad. Sci. USA 78: 3920-3924.

Collmer,A., Berman,P. and Mount,M.S. (1982) Pectate Lyase Regulation and Bacterial Soft-Rot Pathogenesis. p395-442. In M.S. Mount and G.H.Lacy (ed.), Phytopathogenic Prokaryotes, vol. 1. Academic Press, Inc., New York.

Collmer,A. and Keen,N.T. (1986) The Role of Pectic Enzymes in Plant Pathogenesis. Ann. Rev. Phytopathol. 24: 383-409.

Collmer,A., Schoedel,C., Roeder,D.L., Ried,J.L. and Rissler,J.F. (1985) Molecular Cloning in Escherichia coli of Erwinia chrysanthemi Genes Encoding Multiple Forms of Pectate Lyase. J. Bacteriol. 161: 913-920.

Coplin,D.L., Frederick,R.D., Majerczak,D.R. and Haas,E.S. (1986) Molecular Cloning of Virulence Genes from Erwinia stewartii. J. Bacteriol. 168: 619-623.

Crooke,E., Brundage,L., Rice,M. and Wickner,W. (1988) ProOmpA Spontaneously Folds in a Membrane Assembly Competent State which Trigger Factor Stabilizes. EMBO J. 7: 1831-1835.

Crooke,E. and Wickner,W. (1987) Trigger Factor: A Soluble Protein that Folds pro-OmpA into a Membrane-Assembly-Competent Form. Proc. Natl. Acad. Sci. USA 84: 5216-5220.

Cunningham,K., Lill,R., Crooke,E., Rice,M., Moore,K., Wickner,W. and Oliver,D. (1989) SecA Protein, a Peripheral Protein of the Escherichia coli Plasma Membrane, is Essential for the Functional Binding and Translocation of proOmpA. EMBO J. 8: 955-959.

Dahler,G.S., Barras,F. and Keen,N.T. (1990) Cloning of Genes Encoding Extracellular Metalloproteases from Erwinia chrysanthemi EC16. J. Bacteriol. 172: 5803-5815.

Dalrymple,B. and Mattick,J.S. (1987) An Analysis of the Organization and Evolution of Type 4 Fimbrial (MePhe) Subunit Proteins. J. Mol. Evol. 25: 261-269.

Daniels,M.J., Dow,J.M. and Osbourn,A.E. (1988) Molecular Genetics of Pathogenicity in Phytopathogenic Bacteria. Ann. Rev. Phytopathol. 26: 285-312.

Delepelaire,P. and Wandersman,C. (1989) Protein Secretion by Erwinia chrysanthemi. J. Bacteriol. 264: 9083-9089.

Delepelaire,P. and Wandersman,C. (1990) Protein Secretion in Gram Negative Bacteria: The Extracellular Metalloprotease B from Erwinia chrysanthemi Contains a C Terminal Secretion Signal Analogous to that of Escherichia coli α-hemolysin. J. Biol. Chem. 265: 17118-17125.

Devereux,J., Haeberli,P. and Smithies,O. (1984) A Comprehensive Set of Sequence Analysis Programs for the VAX. Nucleic Acids Res. 12: 387-395.

Douglas,C.M., Guidi-Rontani,C. and Collier,R.J. (1987) Exotoxin A of Pseudomonas aeruginosa: Active, Cloned Toxin is Secreted into the Periplasmic Space of Escherichia coli. J. Bacteriol. 169: 4962-4966.

Dow,M.J., Scofield,G., Trafford,K., Turner,P.C. and Daniels,M.J. (1987) A Gene Cluster in Xanthomonas campestris pv. campestris Required for Pathogenicity Controls the Excretion of Polygalacturonate Lyase and Other Enzymes. Physiol. Mol. Plant Pathol. 31: 261-271.

Driks,A., Bryan,R., Shapiro,L. and DeRosier,D.J. (1989) The Organization of the Caulobacter crescentus Flagellar Filament. J. Mol. Biol. 206: 627-636.

Dums,F., Dow,J.M. and Daniels,M.J. (1991) Structural Characterization of Protein Secretion Genes of the Bacterial Phytopathogen Xanthomonas campestris Pathovar campestris: Relatedness to Secretion Systems of Other Gram-Negative Bacteria. Mol. Gen. Genet. In press.

Economou,A., Hamilton,W.D.O., Johnston,A.W.B. and Downie,J.A. (1990) The Rhizobium Nodulation Gene nodO Encodes a Ca++ -Binding Protein that is Exported Without N-terminal Cleavage and is Homologous to Haemolysin and Related Proteins. EMBO J. 9: 349-354.

Eilers,M. and Schatz,G. (1986) Binding of a Specific Ligand Inhibits Import of a Purified Precursor Protein into Mitochondria. Nature 322: 228-232.

Ellard,F.M., Cabello,A. and Salmond,G.P.C. (1989) Bacteriophage λ-Mediated Transposon Mutagenesis of Phytopathogenic and Epiphytic Erwinia Species is Strain Dependent. Mol. Gen. Genet. 218: 491-498.

Elleman,T.C., Hoyne,P.A., McKern,N.M. and Stewart,D.J. (1986) Nucleotide Sequence of the Gene Encoding the Two-Subunit Pilin of Bacteroides nodosus 265. J. Bacteriol. 167: 243-250.

Ellis,J. (1987) Proteins as Molecular Chaperones. Nature 328: 378-379.

Enard,C., Diolez,A. and Expert,D. (1988) Systemic Virulence of Erwinia chrysanthemi 3937 Requires a Functional Iron Assimilation System. J. Bacteriol. 170: 2419-2426.

d'Enfert,C. and Pugsley,A.P. (1987) A Gene Fusion Approach to the Study of Pullulanase Export and Secretion in Escherichia coli. Mol. Microbiol. 1: 159-168.

d'Enfert,C. and Pugsley,A.P. (1989) Klebsiella pneumoniae pulS Gene Encodes an Outer Membrane Lipoprotein Required for Pullulanase Secretion. J. Bacteriol. 171: 3673-3679.

d'Enfert,C., Reyss,I., Wandersman,C. and Pugsley,A.P. (1989) Protein Secretion by Gram-negative Bacteria. J. Biol. Chem. 264:17462-17468.

Engelman,D.M. and Steitz,T.A. (1981) The Spontaneous Insertion of Proteins into and Across Membranes: The Helical Hairpin Hypothesis. Cell 23: 411-422.

Faast,R., Ogierman,M.A., Stroeher,U.H. and Manning,P.A. (1989) Nucleotide Sequence of the Structural Gene, tcpA, for a Major Pilus Subunit of Vibrio cholerae. Gene 85: 227-231.

Finney,K.G., Elleman,T.C. and Stewart,D.J. (1988) Nucleotide Sequence of the Pilin Gene of Bacteroides nodosus 340 (Serogroup D) and Implications for the Relatedness of Serogroups. J. Gen. Microbiol. 134: 575-584.

Focareta,T. and Manning,P.A. (1987) Extracellular Proteins of Vibrio cholerae: Molecular Cloning, Nucleotide Sequence and Characterization of the Deoxyribonuclease (DNase) Together with its Periplasmic Localization in Escherichia coli K-12. Gene 53: 31-40.

Folkhard,W., Marvin,D.A., Watts,T.H. and Paranchych,W. (1981) Structure of Polar Pili from Pseudomonas aeruginosa Strains K and O. J. Mol. Biol. 149: 79-93.

Forbes,K.J. and Perombelon,M.C.M. (1985) Chromosomal Mapping in Erwinia carotovora Subsp. carotovora with the IncP Plasmid R68::Mu. J. Bacteriol. 164: 1110-1116.

Froholm,L.O. and Sletten,K. (1977) Purification and N-Terminal Sequence of a Fimbrial Protein from Moraxella nonliquefaciens. FEBS Lett. 73: 29-32.

Gerlach,J.H., Endicott,J.A., Juranka,P.F., Henderson,G., Sarangi,F., Deuchars,K.L. and Ling,V. (1986) Homology Between P-glycoprotein and a Bacterial Haemolysin Transport Protein Suggests a Model for Multidrug Resistance. Nature 324: 485-489.

van Gijsegem,F. (1986) Analysis of the Pectin-Degrading Enzymes Secreted by Three Strains of Erwinia chrysanthemi. J. Gen. Microbiol. 132: 617-624.

van Gijsegem,F. and Toussaint,A. (1983) In Vivo Cloning of Erwinia carotovora Genes Involved in the Catabolism of Hexuronates. J. Bacteriol. 154:1227-1235.

Gilbert,H.J., Blazek,R., Bullman,H.M.S. and Minton,N.P. (1986) Cloning and Expression of the Erwinia chrysanthemi Asparaginase Gene in Escherichia coli and Erwinia carotovora. J. Gen. Microbiol. 132: 151-160.

Gilkes,N.R., Henrissat,B., Kilburn,D.G., Miller,Jr.,R.C. and Warren,R.A.J. (1991) Domains in Microbial B-1,4-Glycanases: Sequence Conservation, Function, and Enzyme Families. Microbiol. Rev. 55: 303-315.

Gilkes,N.R., Langsford,M.L., Kilburn,D.G., Miller,Jr.,R.C. and Warren,R.A.J. (1984) Mode of Action and Substrate Specificities of Cellulases from Cloned Bacterial Genes. J. Biol. Chem. 259: 10455-10459.

Gilson,L., Mahatny,H.K. and Kolter,R. (1990) Genetic Analysis of an MDR-Like Export System: The Secretion of Colicin V. EMBO J. 9: 3875-3884.

Glaser,P., Sakamoto,H., Bellalou,J., Ullmann,A. and Danchin,A. (1988) Secretion of Cyclolysin, the Calmodulin-Sensitive Adenylate Cyclase - Haemolysin Bifunctional Protein of Bordella pertussis. EMBO J. 7: 3997-4004.

Gobius,K.S. and Pemberton,J.M. (1988) Molecular Cloning, Characterization, and Nucleotide Sequence of an Extracellular Amylase Gene from Aeromonas hydrophila. J. Bacteriol. 170: 1325-1332.

Gold,L. and Stormo,G. (1987) Translational Initiation. In: Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology vol.II. Neidhart,F.C., Ingraham,J.L., Low,K.B., Magasaniti,B., Schaechter,M. and Umbarger,H.E. (eds) American Society for Microbiology, USA., pp 1302-1308

Grindley,J.F., Payton,M.A., van de Pol,H. and Hardy,K.G. (1988) Conversion of Glucose to 2-Keto-L-Gulonate, an Intermediate in L-Ascorbate Synthesis, by a Recombinant Strain of Erwinia citreus. Appl. Environ. Microbiol. 54: 1770-1775.

Guiseppi,A., Cami,B., Aymeric,J.-L., Ball,G. and Creuzet,N. (1988) Homology Between Endoglucanase Z of Erwinia chrysanthemi and Endoglucanases of Bacillus subtilis and alkalophilic Bacillus. Mol. Microbiol. 2: 159-164.

Guzzo,J., Duong,F., Wandersman,C., Murgier,M. and Lazdunski,A. (1991) The Secretion Genes of Pseudomonas aeruginosa Alkaline Protease are Functionally Related to those of Erwinia chrysanthemi Proteases and Escherichia coli α-Haemolysin. Mol. Microbiol. 5: 447-453.

Guzzo,J., Murgier,M., Filloux,A. and Lazdunski,A. (1990) Cloning of the Pseudomonas aeruginosa Alkaline Protease Gene and Secretion of the Protease into the Medium by Escherichia coli. J. Bacteriol. 172: 942-948.

Gygi,D., Nicolet,J., Frey,J., Cross,M., Koronakis,V. and Hughes,C. (1990) Isolation of the Actinobacillus pleuropneumoniae Haemolysin Gene and the Activation and Secretion of the Prohaemolysin by the HlyC, HlyB and HlyD Proteins of Escherichia coli. Mol. Microbiol. 4: 123-128.

Hanahan,D. (1983) Studies on Transformation of Escherichia coli with Plasmids. J. Mol. Biol. 166: 557-580.

Hancock,R.E.W. (1991) Bacterial Outer Membranes: Evolving Concepts. ASM News 57: 175-182.

Harkness,R.E. and Olschlager,T. (1991) The Biology of Colicin M. FEMS Microbiol. Rev. 88: 27-42.

Hartl,F.-U., Ostermann,J., Guiard,B. and Neupert,W. (1987) Successive Translocation Into and Out of the Mitochondrial Matrix: Targeting of Proteins to the Intermembrane Space by a Bipartite Signal Peptide. Cell 51: 1027-1037.

van Haute,E., Joos,H., Maes,M., Warren,G., van Montagu,M. and Schell,J. (1983) Intergeneric Transfer and Exchange Recombination of Restriction Fragments Cloned in pBR322. A Novel Strategy for the Reversed Genetics of the Ti Plasmids of Agrobacterium tumefaciens. EMBO J. 2: 411-417.

He,S.Y. and Collmer,A. (1990) Molecular Cloning, Nucleotide Sequence, and Marker Exchange Mutagenesis of the Exo-Poly- ∝ -D-Galacturonosidase-Encoding pehX Gene of Erwinia chrysanthemi EC16. J. Bacteriol. 172: 4988-4995.

He,S.Y., Lindeberg,M., Chatterjee,A.K. and Collmer,A. (1991a) Cloned Erwinia chrysanthemi out Genes Enable Escherichia coli to Selectively Secrete a Diverse Family of Heterologous Proteins to its Milieu. Proc. Natl. Acad. Sci. USA 88: 1079-1083.

He,S.Y., Schoedel,C., Chatterjee,A.K. and Collmer,A. (1991b) Extracellular Secretion of Pectate Lyase by the Erwinia chrysanthemi Out Pathway is Dependent upon Sec-Mediated Export Across the Inner Membrane. J. Bacteriol. 173: 4310-4317.

von Heijne,G. (1987) Sequence Analysis in Molecular Biology. Academic Press Inc.; San Diego.

Heikinheimo,R., Hemila,H., Pakkanen,R. and Palva,I. (1991) Production of Pectin Methylesterase from Erwinia chrysanthemi B374 in Bacillus subtilis. Appl. Microbiol. Biotechnol. 35: 51-55.

Henrichsen,J. (1983) Twitching Motility. Ann. Rev. Microbiol. 37: 81-93.

Hermondson,M.A., Chen,K.C.S. and Buchanan,T.M. (1978) Neisseria Pili Proteins: Amino-terminal Amino Acid Sequences and Identification of an Unusual Amino Acid. Biochemistry 17: 442-445.

Higgins,C.F., Hiles,I.D., Salmond,G.P.C., Gill,D.R., Downie,J.A., Evans,I.J., Holland,I.B., Gray,L., Buckel,S.D., Bell,A.W. and Hermondson,M.A. (1986) A Family of Related ATP-Binding Subunits Coupled to Many Distinct Biological Processes in Bacteria. Nature 323: 448-450.

Hinton,J.C.D. (1986) Development and Application of Molecular Genetic Techniques in Erwinia carotovora subsp. carotovora. PhD thesis, Warwick University.

Hinton,J.C.D., Gill,D.R., Lalo,D., Plastow,G.S. and Salmond,G.P.C. (1990) Sequence of the peh Gene of Erwinia carotovora: Homology Between Erwinia and Plant Enzymes. Mol. Microbiol. 4: 1029-1036.

Hinton,J.C.D., Perombelon,M.C.M. and Salmond,G.P.C. (1985a) Efficient Transformation of Erwinia carotovora subsp. carotovora and E. carotovora subsp. atroseptica. J. Bacteriol. 161:786-788.

Hinton,J.C.D., Perombelon,M.C.M. and Salmond,G.P.C. (1985b) Nonsense Suppressor Mutants of Erwinia carotovora subsp. carotovora. FEMS Microbiol. Lett. 28: 103-106.

Hinton,J.C.D. and Salmond,G.P.C. (1987) Use of TnphoA to Enrich for Extracellular Enzyme Mutants of Erwinia carotovora Subspecies carotovora. Mol. Microbiol. 1: 381-386.

Hinton,J.C.D., Sidebotham,J.M., Gill,D.R. and Salmond,G.P.C. (1989a) Extracellular and Periplasmic Isoenzymes of Pectate Lyase from Erwinia carotovora Subspecies carotovora Belong to Different Gene Families. Mol. Microbiol. 3: 1785-1795.

Hinton,J.C.D., Sidebotham,J.M., Hyman,L.J., Perombelon,M.C.M. and Salmond,G.P.C. (1989b) Isolation and Characterization of Transposon-Induced Mutants of Erwinia carotovora Subsp. atroseptica Exhibiting Reduced Virulence. Mol. Gen. Genet. 217: 141-148.

Hirst,T.R. and Holmgren,J. (1987) Conformation of Protein Secreted Across Bacterial Outer Membranes: A Study of Enterotoxin Translocation from Vibrio cholerae. Proc. Natl. Acad. Sci. USA 84: 7418-7422.

Hohn,B. and Collins,J. (1980) A Small Cosmid for Efficient Cloning of Large DNA Fragments. Gene 11: 291-298.

Holland,B., Mackman,N. and Nicaud,J.-M. (1986) Secretion of Proteins from Bacteria. Biotechnol. 4: 427-431.

Holland,I.B., Kenny,B. and Blight,M. (1990) Haemolysin Secretion from E. coli. Biochimie 72: 131-141

Holmegren,A. and Branden,C.-I. (1989) Crystal Structure of Chaperone Protein PapD Reveals an Immunoglobulin Fold. Nature 342: 248-252.

Homma,M., Fujita,H., Yamaguchi,S. and Iino,T. (1984) Excretion of Unassembled Flagellin by Salmonella typhimurium Mutants Deficient in Hook-Associated Proteins. J. Bacteriol. 159: 1056-1059.

Howard,S.P. and Buckley,J.T. (1986) Molecular Cloning and Expression in Escherichia coli of the Structural Gene for the Hemolytic Toxin Aerolysin from Aeromonas Hydrophila. Mol. Gen. Genet. 204: 289-295.

Howard,S.P., Cavard,D. and Lazdunski,C. (1991) Phospholipase-A-Independent Damage Caused by the Colicin A Lysis Protein During its Assembly into the Inner and Outer Membranes of Escherichia coli. J. Gen. Microbiol. 137: 81-89.

Huang,J. and Schell,M.A. (1990a) DNA Sequence Analysis of pglA and Mechanism of Export of its Polygalacturonase Product from Pseudomonas solanacearum. J. Bacteriol. 172: 3879-3887.

Huang,J., and Schell,M.A. (1990b) Evidence that Extracellular Export of the Endoglucanase Encoded by egl of Pseudomonas solanacearum Occurs by a Two-Step Process Involving a Lipoprotein Intermediate. J. Biol. Chem. 265: 11628-11632.

Hugouvieux-Cotte-Pattat,N., Reverchon,S., Condemine,G. and Robert-Baudouy,J. (1986) Regulatory Mutations Affecting the Synthesis of Pectate Lyase in Erwinia chrysanthemi. J. Gen. Microbiol. 132: 2099-2106.

Ishimaru,C.A., Klos,E.J. and Brubaker,R.R. (1988) Multiple Antibiotic Production by Erwinia herbicola. Phytopathology 78: 746-750.

Jayswal,R.K., Bressan,R.A. and Handa,A.K. (1984) Mutagenesis of Erwinia carotovora Subsp. carotovora with Bacteriophage Mu d1(Ap^r lac cts62): Construction of his-lac Gene Fusions. J. Bacteriol. 158: 764-766.

Ji,J., Hugouvieux-Cotte-Pattat,N. and Robert-Baudouy,J. (1987) Use of Mu-lac Insertions to Study the Secretion of Pectate Lyases by Erwinia chrysanthemi. J. Gen. Microbiol. 133: 793-802.

Ji,J., Hugouvieux-Cotte-Pattat,N. and Robert-Baudouy,J. (1989) Molecular Cloning of the outJ Gene Involved in Pectate Lyase Secretion by Erwinia chrysanthemi. Mol. Microbiol. 3: 285-293.

Jiang,B. and Howard,S.P. (1991) Mutagenesis and Isolation of Aeromonas hydrophila Genes Which are Required for Extracellular Secretion. J. Bacteriol. 173: 1241-1249.

Kallio,P., Simonen,M., Palva,I. and Sarvas,M. (1986) Synthesis of OmpA Protein of Escherichia coli K12 in Bacillus subtilis. J. Gen. Microbiol. 132: 677-687.

Keen,N.T., Dahlbeck,D., Staskawicz,B. and Belser,W. (1984) Molecular Cloning of Pectate Lyase Genes from Erwinia chrysanthemi and Their Expression in Escherichia coli. J. Bacteriol. 159: 825-831.

Kieser (1984) Plasmid Isolation by Alkaline Lysis (Streptomyces or E. coli): Procedure 2. In: Genetic Manipulation of Streptomyces. A Laboratory Manual. Hopwood,D.A., Bibb,M.J., Chater,K.F., Kieser,T., Bruton,C.J., Kieser,H.M., Lydiate,D.J., Smith,C.P., Ward,J.M. and Schrempf,H. (eds) The John Innes Foundation; Norwich, pp 84-92.

Klauser,T., Pohlner,J. and Meyer,T.F. (1990) Extracellular Transport of Cholera Toxin B Subunit Using Neisseria IgA Protease β-Domain: Conformation-Dependent Outer Membrane Translocation. EMBO J. 9: 1991-1999.

Kontinen,V.P., Saris,P. and Sarvas,M. (1991) A Gene (prsA) of Bacillus subtilis Involved in a Novel Late Stage of Protein Export. Mol. Microbiol. 5: 1273-1283.

Kontinen,V.P. and Sarvas,M. (1988) Mutants of Bacillus subtilis Defective in Protein Export. J. Gen. Microbiol. 134: 2333-2344.

Koomey,M., Bergstrom,S., Blake,M. and Swanson,J. (1991) Pilin Expression and Processing in Pilus Mutants of Neisseria gonorrhoeae: Critical Role of Gly-1 in Assembly. Mol. Microbiol. 5: 279-287.

Kornacker,M.G., and Pugsley,A.P. (1989) Molecular Characterization of pulA and Its Product, Pullulanase, a Secreted Enzyme of Klebsiella pneumoniae UNF5023. Mol. Microbiol. 4: 73-85.

Kornacker,M.G. and Pugsley,A.P. (1990) The Normally Periplasmic Enzyme B-lactamase is Specifically and Efficiently Translocated Through the Escherichia coli Outer Membrane when it is Fused to the Cell-Surface Enzyme Pullulanase. Mol. Microbiol. 4: 1101-1100.

Koronakis,V., Cross,M., Senior,B., Koronakis,E. and Hughes,C. (1987) The Secreted Hemolysins of Proteus mirabilis, Proteus vulgaris, and Morganella morganii are Genetically Related to Each Other and to the Alpha-Hemolysin of Escherichia coli. J. Bacteriol. 169: 1509-1515.

Koronakis,V., Koronakis,E. and Hughes,C. (1988) Comparison of the Haemolysin Secretion Protein HlyB from Proteus vulgaris and Escherichia coli; Site-Directed Mutagenesis Causing Impairment of Export Function. Mol. Gen. Genet. 213: 551-555.

Koronakis,V., Koronakis,E. and Hughes,C. (1989) Isolation and Analysis of the C-Terminal Signal Directing Export of Escherichia coli Hemolysin Protein Across Both Bacterial Membranes. EMBO J. 8: 595-605.

Kotoujansky,A. (1987) Molecular Genetics of Pathogenesis by Soft-Rot Erwinias. Ann. Rev. Phytopathol. 25: 405-430.

Kotoujansky,A., Diolez,A., Boccara,M., Bertheau,Y., Andro,T. and Coleno,A. (1985) Molecular Cloning of Erwinia chrysanthemi Pectinase and Cellulase Structural Genes. EMBO J. 4: 781-785.

Kuhn,A., Rohrer,J. and Gallusser,A. (1990a) Bacteriophages M13 and Pf3 Tell Us How Proteins Insert into the Membrane. J. Struct. Biol. 104: 38-43.

Kuhn,A., Zhu,H.-Y. and Dalbey,R.E. (1990b) Efficient Translocation of Positively Charged Residues of M13 Procoat Protein Across the Membrane Excludes Electrophoresis as the Primary Force for Membrane Insertion. EMBO J. 9: 2385-2389.

Kuldau,G.A., De Vos,G., Owen,J., McCaffrey,G. and Zambryski,P. (1990) The virB Operon of Agrobacterium tumefaciens pTiC58 Encodes 11 Open Reading Frames. Mol. Gen. Genet. 221: 256-266.

Kumar,R., Ghosh,A. and Ghosh,B.K. (1983) Alkaline Phosphatase Secretion-Negative Mutant of Bacillus licheniformis 749/C. J. Bacteriol. 154: 946-954.

Kusukawa,N., Yura,T., Ueguchi,C., Akiyama,Y. and Ito,K. (1989) Effects of Mutations in Heat-Shock Genes groES and groEL on Protein Export in Escherichia coli. EMBO J. 8: 3517-3521.

Kyte,J. and Doolittle,R.F. (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. J. Mol. Biol. 157: 105-132.

Laemmli,U.K. (1970) Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4. Nature 227: 680-685.

Lazdunski,A., Guzzo,J., Filloux,A., Bally,M. and Murgier,M. (1990) Secretion of Extracellular Proteins by Pseudomonas aeruginosa. Biochimie 72: 147-156.

Lecker,S., Lill,R., Ziegelhoffer,T., Georgopoulos,C., Bassford,Jr.,P.J., Kumamoto,C.A. and Wickner,W. (1989) Three Pure Chaperone Proteins of Escherichia coli - SecB, Trigger Factor and GroEL - Form Soluble Complexes with Precursor Proteins in vitro. EMBO J. 8: 2703-2709.

Lei,S.-P., Lin,H.-C., Heffernan,L. and Wilcox,G. (1985) Evidence that Polygalacturonase is a Virulence Determinant in Erwinia carotovora. J. Bacteriol. 164: 831-835.

Lelliott,R.A. and Dickey,R.S. (1984) Genus VII: Erwinia. In Krieg,N.R. and Holt,J.G. (eds) Bergey's Manual of Systematic Bacteriology, vol. 1. Williams and Wilkins, Baltimore, London, pp 469-476.

Letoffe,S., Delepelaire,P. and Wandersman,C. (1990) Protease Secretion by Erwinia chrysanthemi: the Specific Secretion Functions are Analogous to those of Escherichia coli ∝-haemolysin. EMBO J. 9: 1375-1382.

Letoffe,S., Delepelaire,P. and Wandersman,C. (1991) Cloning and Expression in Escherichia coli of the Serratia marcescens Metalloprotease Gene: Secretion of the Protease from E. coli in the Presence of the Erwinia chrysanthemi Protease Secretion Functions. J. Bacteriol. 173: 2160-2166.

Lill,R., Cunningham,K., Brundage,L.A., Ito,K., Oliver,D. and Wickner,W. (1989) SecA Protein Hydrolyzes ATP and is an Essential Component of the Protein Translocation ATPase of Escherichia coli. EMBO J. 8: 961-966.

Lindberg,F., Tennent,J.M., Hultgren,S.J., Lund,B. and Normark,S. (1989) PapD, a Periplasmic Transport Protein in P-Pilus Biogenesis. J. Bacteriol. 171: 6052-6058.

Lory,S., Strom,M.S. and Johnson,K. (1988) Expression and Secretion of the Cloned Pseudomonas aeruginosa Exotoxin A by Escherichia coli. J. Bacteriol. 170: 714-719.

Luiten,R.G.M., Putterman,D.G., Schoenmakers,J.G.G., Konings,R.N.H. and Day,L.A. (1985) Nucleotide Sequence of the Genome of Pf3, an IncP-1 Plasmid-Specific Filamentous Bacteriophage of Pseudomonas aeruginosa. J. Virol. 56: 268-276.

Lugtenberg,B., and van Alphen,L. (1983) Molecular Architecture and Functioning of the Outer Membrane of E. coli and Other G- Bacteria. Biochim. Biophys. Acta 737: 51-155

Mackman,N., Baker,K., Gray,L., Haigh,R., Nicaud,J.-M. and Holland,I.B. (1987) Release of a Chimeric Protein into the Medium from Escherichia coli Using the C-terminal Secretion Signal of Haemolysin. EMBO J. 6: 2835-2841.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Manulis,S., Kobayashi,D.Y. and Keen,N.T. (1988) Molecular Cloning and Sequencing of a Pectate Lyase Gene from Yersinia pseudotuberculosis. J. Bacteriol. 170: 1825-1830.

Marrs,C.F., Schoolnik,G., Koomey,J.M., Hardy,J., Rothbard,J. and Falkow,S. (1985) Cloning and Sequencing of a Moraxella bovis Pilin Gene. J. Bacteriol. 163: 132-139.

Mattick,J.S., Anderson,B.J., Cox,P.T., Dalrymple,B.P., Bills,M.M., Hobbs,M. and Egerton,J.R. (1991) Gene Sequences and Comparison of the Fimbrial Subunits Representative of Bacteroides nodosus Serotypes A to I: Class I and Class II Strains. Mol. Microbiol. 5: 561-573.

McEvoy,J.L., Murata,H. and Chatterjee,A.K. (1990) Molecular Cloning and Characterization of an Erwinia carotovora Subsp. carotovora Pectin Lyase Gene that Responds to DNA-Damaging Agents. J. Bacteriol. 172: 3284-3289.

Meyer,D.I. (1988) Preprotein Conformation: the Year's Major Theme in Translocation Studies. TIBS 13: 471-474.

Meyer,T.F., Billyard,E., Haas,R., Storzbach,S. and So,M. (1984) Pilus Genes of Neisseria gonorrhoeae: Chromosomal Organization and DNA Sequence. Proc. Natl. Acad. Sci. USA 81: 6110-6114.

Meyer,D.I. and Dobberstein,B. (1980) Identification and Characterization of a Membrane Component Essential for the Translocation of Nascent Proteins Across the Membrane of the Endoplasmic Reticulum. J. Cell Biol. 87: 503-508.

Meyer,D.I., Krause,E. and Dobberstein,B. (1982) Secretory Protein Translocation Across Membranes - the Role of the 'Docking Protein'. Nature 297: 647-650.

Michiels,T., Vanooteghem,J.-C., Lambert de Rouvroit,C., China,B., Gustin,A., Boudry,P. and Cornelis,G.R. (1991) Analysis of virC, an Operon Involved in the Secretion of Yop Proteins by Yersinia enterocolitica. J. Bacteriol. 173: 4994-5009.

Miller,J.H. (1972) In Experiments in Molecular Genetics. Cold Spring Harbor, Cold Spring Harbor Laboratory Press; New York pp. 412-414.

Milstein,C., Brownlee,G.G., Harrison,T.M. and Mathews,M.B. (1972) A Possible Precursor of Immunoglobulin Light Chains. Nature New Biology 239:117-120.

Misawa,N., Yamano,S. and Ikenaga,H. (1991) Production of B-Carotene in Zymomonas mobilis and Agrobacterium tumefaciens by Introduction of the Biosynthesis Genes from Erwinia uredovora. Appl. Environ. Microbiol. 57: 1847-1849.

Miyazaki,H., Yanagida,N., Horinouchi,S. and Beppu,T. (1989) Characterization of the Precursor of Serratia marcescens Serine Protease and COOH-Terminal Processing of the Precursor During its Excretion Through the Outer Membrane of Escherichia coli. J. Bacteriol. 171: 6566-6572.

Mohan,S., Aghion,J., Guillen,N. and Dubnau,D. (1989) Molecular Cloning and Characterization of comC, a Late Competence Gene of Bacillus subtilis. J. Bacteriol. 171: 6043-6051.

Murata,H., Fons,M., Chatterjee,A., Collmer,A. and Chatterjee,A.K. (1990) Characterization of Transposon Insertion Out- Mutants of Erwinia carotovora subsp. carotovora Defective in Enzyme Export and of a DNA Segment that Complements out Mutations in E. carotovora subsp. carotovora, E. carotovora subsp. atroseptica, and Erwinia chrysanthemi. J. Bacteriol. 172: 2970-2978.

Murphy,G. and Kavanagh,T. (1988) Speeding-Up the Sequencing of Double-Stranded DNA. Nucleic Acids Res. 16: 5198-5201.

Nakahama,K., Yoshimura,K., Marumoto,R., Kikuchi,M., Lee,I.S., Hase,T. and Matsubara,H. (1986) Cloning and Sequencing of Serratia Protease Gene. Nucleic Acids Res. 14: 5843-5855.

Nagley,P. and Devenish,R.J. (1989) Leading Organellar Proteins Along New Pathways: the Relocation of Mitochondrial and Chloroplast Genes to the Nucleus. TIBS 14: 31-35.

Nambudripad,R., Stark,W., Opella,S.J. and Makowski,L. (1991) Membrane-Mediated Assembly of Filamentous Bacteriophage Pf1 Coat Protein. Science 252: 1305-1308

Neu,H.C. and Heppel,L.A. (1965) The Release of Enzymes from Escherichia coli Osmotic Shock and During the Formation of Spheroplasts. J. Biol. Chem. 240: 3685-3692.

Nicaud,J.-M., Breton,A., Younes,G. and Guespin-Michel,J. (1984) Mutants of Myxococcus xanthus Impaired in Protein Secretion: An Approach to Study of a Secretory Mechanism. Appl. Microbial. Biotechnol. 26: 344-350.

Novick,P, Ferro,S. and Schekman,R. (1981) Order of Events in the Yeast Secretory Pathway. Cell 25: 461-469.

Nunn,D., Bergman,S. and Lory,S. (1990) Products of Three Accessory Genes, pilB, pilC, and pilD, are Required for Biogenesis of Pseudomonas aeruginosa Pili. J. Bacteriol. 172: 2911-2919.

Orvos,D.R., Lacy,G.H. and Cairns,Jr.,J. (1990) Genetically Engineered Erwinia carotovora: Survival, Intraspecific Competition, and Effects upon Selected Bacterial Genera. Appl. Environ. Microbiol. 56: 1689-1694.

Osborn,M.J. and Munson,R. (1974) Separation of the Inner (Cytoplasmic) and Outer Membranes of Gram-Negative Bacteria. Methods in Enzymology 31: 642-653.

Ottow,J.C.G. (1975) Ecology, Physiology, and Genetics of Fimbriae and Pili. Ann. Rev. Microbiol. 29: 79-108.

Palade,G. (1975) Intracellular Aspects of the Process of Protein Synthesis. Science 189: 347-358.

Parker,W.L., Rathnum,M.L., Wells,Jr.,J.S., Trejo,W.H., Principe,P.A. and Sykes,R.B. (1982) SQ 27,860, a Simple Carbapenem Produced by Species of Serratia and Erwinia. J. Antibiotics 35: 653-660.

Pasloske,B.L., Sastry,P.A., Finlay,B.B. and Paranchych,W. (1988) Two Unusual Pilin Sequences from Different Isolates of Pseudomonas aeruginosa. J. Bacteriol. 170: 3738-3741.

Pearson, G.D.N. and Mekalanos, J.J. (1982) Molecular Cloning of Vibrio cholerae Enterotoxin Genes in Escherichia coli K-12. Proc. Natl. Acad. Sci. USA. 79: 2976-2980.

Peeters,B.P.H., Peters,R.M., Schoenmakers,J.G.G. and Konings,R.N.H. (1985) Nucleotide Sequence and Genetic Organization of the Genome of the N-Specific Filamentous Bacteriophage IKe. Comparison with the Genome of the F-Specific Filamentous Phages M13, fd and f1. J. Mol. Biol. 181: 27-39.

Perombelon,M.C.M. (1987) Pathogenesis by Pectolytic Erwinias. In Civerolo,E.L., Collmer,A., Gillaspie,A.G. and Davis,R.E. (eds) Proc. 6th Int. Conf. Plant Path. Bact., Maryland. Martinus Nijhoff, Dordrecht, pp 109-120.

Perombelon,M.C.M. and Kelman,A. (1980) Ecology of the Soft Rot Erwinias. Ann. Rev. Phytopathol. 18: 361-387.

Perombelon,M.C.M. and Kelman,A. (1987) Blackleg and Other Potato Diseases Caused by Soft Rot Erwinias: Proposal for Revision of Terminology. Phytopathology 71: 283-285.

Pfanner,N., Hartl,F.-U. and Neupert,W. (1988) Import of Proteins into Mitochondria: a Multi-Step Process. Eur. J. Biochem. 175: 205-212.

Plastow,G.S. (1988) Molecular Cloning and Nucleotide Sequence of the Pectin Methyl Esterase Gene of Erwinia chrysanthemi B374. Mol. Microbiol. 2: 247-254.

Plastow,G.S., Border,P.M., Hinton,J.C.D. and Salmond,G.P.C. (1986) Molecular Cloning of Pectinase Genes from Erwinia carotovora Subspecies carotovora (Strain SCRI193). Symbiosis 2: 115-122.

Pohlner,J., Halter,R., Beyreuther,K. and Meyer,T.F. (1987) Gene Structure and Extracellular Secretion of Neisseria gonorrhoeae IgA Protease. Nature 325: 458-462.

Potts,W.J. and Saunder,J.R. (1988) Nucleotide Sequence of the Structural Gene for Class I Pilin from Neisseria meningitidis: Homologies with the pilE Locus of Neisseria gonorrhoeae. Mol. Microbiol. 2: 647-653.

Pugsley,A.P. (1988) Protein Secretion Across the Outer Membrane of Gram-Negative Bacteria. In Das,R.C. and Robbins,P.W. (eds). Protein Transfer and Organelle Biogenesis; Academic Press, Inc., Orlando, Fla, pp 607-652.

Pugsley,A.P., d'Enfert,C., Reyss,I. and Kornacker,M.G. (1990a) Genetics of Extracellular Protein Secretion by Gram-Negative Bacteria. Annu. Rev. Genet. 24: 67-90.

Pugsley,A.P., Kornacker,M.G. and Poquet,I. (1991a) The General Protein-Export Pathway is Directly Required for Extracellular Pullulanase Secretion in Escherichia coli K12. Mol. Microbiol. 5(2): 343-352.

Pugsley,A.P., Kornacker,M.G. and Ryter,A. (1990b) Analysis of the Subcellular Location of Pullulanase Produced by Escherichia coli Carrying the pulA Gene from Klebsiella pneumoniae Strain UNF5023. Mol. Microbiol. 4: 59-72.

Pugsley, A.P., Poquet,I. and Kornacker,M.G. (1991b) Two Distinct Steps in Pullulanase Secretion by Escherichia coli K12. Mol. Microbiol. 5(4): 865-873.

Pugsley,A.P. and Reyss,I. (1990) Five Genes at the 3' End of the Klebsiella pneumoniae pulC Operon are Required for Pullulanase Secretion. Mol. Microbiol. 4: 365-379.

Pugsley,A.P. and Schwartz,M. (1985) Export and Secretion of Proteins by Bacteria. FEMS Microbiol. Rev. 32: 3-38.

Py,B., Salmond,G.P.C., Chippaux,M. and Barras,F. (1991) Secretion of Cellulases in Erwinia chrysanthemi and E. carotovora is Species-Specific. FEMS Microbiol. Lett. 79: 315-322.

Queen,C. and Korn,L.J. (1984) A Comprehensive Sequence Analysis Programme for the IBM Personal Computer. Nucl. Acids Res. 12: 581-599.

Randall,L.L., Hardy,S.J.S. and Thom, J.R. (1987) Export of Protein: A Biochemical View. Ann. Rev. Microbiol. 41: 507-541.

Randall,L.L., Topping,T.B. and Hardy,S.J.S. (1990) No Specific Recognition of Leader Peptide by SecB, a Chaperone Involved in Protein Export. Science 248: 860-863.

Reverchon,S. and Robert-Baudouy,J. (1987) Regulation of Expression of Pectate Lyase Genes pelA, pelD and pelE in Erwinia chrysanthemi. J. Bacteriol. 169: 2417-2423.

Reyss,I. and Pugsley,A.P. (1990) Five Additional Genes in the pulC-O Operon of the Gram-Negative Bacterium Klebsiella oxytoca UNF5023 which are Required for Pullulanase Secretion. Mol. Gen. Genet. 222: 176-184.

Ried,J.L. and Collmer,A. (1988) Construction and Characterization of an Erwinia chrysanthemi Mutant with Directed Deletions in All of the Pectate Lyase Structural Genes. Mol. Plant-Microbe Interact. 1: 32-38.

Riordan,J.R., Rommens,J.M., Kerem,B.S., Alon,N., Rozmahel,R., Grzelczak,Z., Zielenski,J., Lok,S., Plavsic,N., Chou,J.L., Drum,M.L., Iannuzzi,M.C., Collins,F.S., Tsui,L.C. Identification of the Cystic Fibrosis Gene-Cloning and Complementation of Complementary-DNA. (1989) Science 245: 1066-1073.

Roeder,D.L. and Collmer,A. (1985) Marker-Exchange Mutagenesis of a Pectate Lyase Isozyme Gene in Erwinia chrysanthemi. J. Bacteriol. 164: 51-56.

Russel,M. (1991) Filamentous Phage Assembly. Mol. Microbiol. 5: 1607-1613.

Saarilahti,H.T., Heino,P., Pakkanen,R., Kalkkinen,N., Palva,I. and Palva,E.T. (1990a) Structural Analysis of the pehA Gene and Characterization of its Protein Product, Endopolygalacturonase, of Erwinia carotovora Subspecies carotovora. Mol. Microbiol. 4: 1037-1044.

Saarilahti,H.T., Henrissat,B. and Palva,E.T. (1990b) CelS: a Novel Endoglucanase Identified from Erwinia carotovora Subsp. carotovora. Gene 90: 9-14.

Saier,M.H.,Jr., Werner,P.K. and Muller,M. (1989) Insertion of Proteins into Bacterial Membranes: Mechanism,Characteristics, and Comparisons with the Eucaryotic Process. Microbiol. Rev. 53: 333-366.

Salmond,G.P.C., Hinton,J.C.D., Gill,D.R. and Perombelon,M.C.M. (1986) Transport Mutagenesis of Erwinia Using Phage λ Vectors. Mol. Gen. Genet. 203: 524-528.

Sambrook,J. and Gething,M.-J. (1989) Chaperones,Paperones. Nature 342: 224-225.

Sancar,A., Hack,A.M. and Rupp,W.D. (1979) Simple Method for Identification of Plasmid-Coded Proteins. J. Bacteriol. 137: 692-693.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA Sequencing with Chain-Terminating Inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schatz,P.J., Bieker,K.L., Ottemann,K.M., Silhavy,T.J. and Beckwith,J. (1991) One of Three Transmembrane Stretches is Sufficient for the Functioning of the SecE Protein, a Membrane Component of the E.coli Secretion Machinery. EMBO J. 10: 1749-1757.

Schiebel,E., Driessen,A.J.M., Hartl,F.-U. and Wickner,W. (1991) $\Delta \mu H+$ and ATP Function at Different Steps of the Catalytic Cycle of Preprotein Translocase. Cell 64: 927-939.

Schoonejans,E., Expert,D. and Toussaint,A. (1987) Characterization and Virulence Properties of Erwinia chrysanthemi Lipopolysaccharide-Defective, øEC2-Resistant Mutants. J. Bacteriol. 169: 4011-4017.

Seifert,H.S., Ajioka,R.S., Paruchuri,D., Heffron,F. and So,M. (1990) Shuttle Mutagenesis of Neisseria gonorrhoeae: Pilin Null Mutations Lower DNA Transformation Competence. J. Bacteriol. 172: 40-46.

Shirae,H., Yokozeki,K. and Kubota,K. (1988) Enzymatic Production of Ribavirin from Orotidine by Erwinia carotovora AJ 2992. Agric. Bio. Chem. 52: 1499-1504.

Shirasu,K., Morel,P. and Kado,C.I. (1990) Characterization of the virB Operon of an Agrobacterium tumefaciens Ti Plasmid: Nucleotide Sequence and Protein Analysis. Mol. Microbiol. 4: 1153-1163.

Shon,K.-J., Kim,Y., Colnago,L.A. and Opella,S.J. (1991) NMR Studies of the Structure and Dynamics of Membrane-Bound Bacteriophage Pf1 Coat Protein. Science 252: 1303-1305.

Singer,S.J., Maher,P.A. and Yaffe,M.P. (1987) On the Translocation of Proteins Across Membranes. Proc. Natl. Sci. USA 84: 1015-1019.

Spinnler,H.E. and Djian,A. (1991) Bioconversion of Amino Acids into Flavouring Alcohols and Esters by Erwinia carotovora Subsp. atroseptica. Appl. Microbiol. Biotechnol. 35: 264-269.

Staden,R. (1978) Further Procedures for Sequence Analysis by Computer. Nucl. Acids Res. 5: 1013-1016.

Stanier,R.Y., Adelberg,E.A. and Ingraham,J.L. (1980) General Microbiology (4th Edition) Macmillan Press Ltd. p409.

Sternberg,N. and Weisberg,R. (1975) Packaging of Prophage and Host DNA by Coliphage λ. Nature 256: 97-103.

Strathdee,C.A. and Lo,R.Y.C. (1989) Cloning, Nucleotide Sequence, and Characterization of Genes Encoding the Secretion Function of the Pasteurella haemolytica Leukotoxin Determinant. J. Bacteriol. 171: 916-928.

Strom,M.K. and Lory,S. (1986) Cloning and Expression of the Pilin Gene of Pseudomonas aeruginosa PAK in Escherichia coli. J. Bacteriol. 165: 367-372.

Struyve,M., Moons,M. and Tommassen,J. (1991) Carboxy-terminal Phenylalanine is Essential for the Correct Assembly of a Bacterial Outer Membrane Protein. J. Mol. Biol. 218: 141-148.

Swidersky,U.E., Hoffschulte,H.K. and Muller,M. (1990) Determinants of Membrane-Targeting and Transmembrane Translocation During Bacterial Protein Export. EMBO J. 9: 1777-1785.

Tabor,S. and Richardson,C.C. (1985) A Bacteriophage T7 RNA Polymerase/Promoter System for Controlled Exclusive Expression of Specific Genes. Proc. Natl. Acad. Sci. USA 82: 1074-1078.

Tamaki,S.J., Gold,S., Robeson,M., Manulis,S. and Keen,N.T. (1988) Structure and Organization of the pel Genes from Erwinia chrysanthemi EC16. J. Bacteriol. 170: 3468-3478.

Thurn,K.K. and Chatterjee,A.K. (1985) Single-Site Chromosomal Tn5 Insertions Affect the Export of Pectolytic and Cellulolytic Enzymes in Erwinia chrysanthemi EC16. App. Environ. Microbiol. 50: 894-898.

Trollinger,D., Berry,S., Belser,W. and Keen,N.T. (1989) Cloning and Characterization of a Pectate Lyase Gene from Erwinia carotovora EC153. Mol. Plant Microb. Interact. 2: 17-25.

Tsuyumu,S. (1979) 'Self-Catabolite Repression' of Pectate Lyase in Erwinia carotovora. J. Bacteriol. 137: 1035-1036.

Tsuyumu,S. and Chatterjee,A.K. (1984) Pectin Lyase Production in Erwinia chrysanthemi and Other Soft-Rot Erwinia Species. Physiol. Plant Pathol. 24: 291-302.

Uphoff,T.S. and Welch,R.A. (1990) Nucleotide Sequencing of the Proteus mirabilis Calcium-Independent Hemolysin Genes (hpmA and hpmB) Reveals Sequence Similarity with the Serratia marcescens Hemolysin Genes (shlA and shlB). J. Bacteriol. 172: 1206-1216.

Vollenweider,H.J., Fiandt,M., Rosenvold,E.C. and Szybalski,W. (1980) Packaging of Plasmid DNA Containing the Cohesive Ends of Coliphage Lambda. Gene 9: 171-174.

Walderich,B. and Holtje,J.-V. (1989) Specific Localization of the Lysis Protein of Bacteriophage MS2 in Membrane Adhesion Sites of Escherichia coli. J. Bacteriol. 171: 3331-3336.

Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) Distantly Related Sequences in the Alpha and Beta Subunits of ATP Synthase, Myosin, Kinases and Other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold. EMBO J. 1: 945-951.

Walter,P. and Blobel,G. (1980) Purification of a Membrane-Associated Protein Complex Required for Protein Translocation Across the Endoplasmic Reticulum. Proc. Natl. Acad. Sci. USA 77: 7112-7116.

Walter,P. and Blobel,G. (1981) Translocation of Proteins Across the Endoplasmic Reticulum III. Signal Recognition Protein (SRP) Causes Signal Sequence-Dependent and Site-Specific Arrest of Chain Elongation that is Released by Microsomal Membranes. J. Cell Biol. 91: 557-561.

Wandersman,C. (1989) Secretion, Processing and Activation of Bacterial Extracellular Proteases. Mol. Microbiol. 3: 1825-1831.

Wandersman,C. and Delepelaire,P. (1990) TolC, an Escherichia coli Outer Membrane Protein Required for Hemolysin Secretion. Proc. Natl. Acad. Sci. USA 87: 4776-4780.

Wandersman,C., Delepelaire,P. and Letoffe,S. (1990) Secretion Processing and Activation of Erwinia chrysanthemi Proteases. Biochimie 72: 143-146

Wandersman,C., Delepelaire,P., Letoffe,S. and Schwartz,M. (1987) Characterization of Erwinia chrysanthemi Extracellular Proteases: Cloning and Expression of the Protease Genes in Escherichia coli. J. Bacteriol. 169: 5046-5063.

Ward,J.E., Akiyoshi,D.E., Regier,D., Datta,A., Gordon,M.P. and Nester,E.W. (1988) Characterization of the virB Operon from an Agrobacterium tumefaciens Ti Plasmid. J. Biol. Chem. 263: 5804-5814.

Ward,J.E.,Jr., Dale,E.M., Nester,E.W. and Binns,A.N. (1990) Identification of a VirB10 Protein Aggregate in the Inner Membrane of Agrobacterium tumefaciens. J. Bacteriol. 172: 5200-5210.

Watanabe,M. and Blobel,G. (1989) SecB Functions as a Cytosolic Signal Recognition Factor for Protein Export in E. coli. Cell 58: 695-705.

Wells,J.M., Sheng,W.-S., Ceponis,M.J. and Chen,T.A. (1987) Isolation and Characterization of Strains of Erwinia ananas from Honeydew Melons. Phytopathology 77: 511-514.

van Wezenbeek,P.M.G.F., Hulsebos,T.J.M. and Schoenmakers,J.G.G. (1980) Nucleotide Sequence of the Filamentous Bacteriophage M13 Genome: Comparison with Phage fd. Gene 11: 129-148.

Whitchurch,C.B., Hobbs,M., Livingstone,S.P., Krishnapillai,V. and Mattick,J.S. (1990) Characterization of a Pseudomonas aeruginosa Twitching Motility Gene and Evidence for a Specialised Protein Export System Widespread in Eubacteria. Gene 101: 33-44.

White,F.F., Klee,H.J. and Nester,E.W. (1983) In Vivo Packaging of Cosmids in Transposon-Mediated Mutagenesis. J. Bacteriol. 153: 1075-1078.

Wickner,W. (1989) Secretion and Membrane Assembly. TIBS 14: 280-283.

Willis,J.W., Engwall,J.K. and Chatterjee,A.K. (1987) Cloning of Genes for Erwinia carotovora Subsp. carotovora Pectolytic Enzymes and Further Characterization of the Polygalacturonases. Mol. Plant Pathol. 77: 1199-1205.

Witte,A., Blasi,U., Halfmann,G., Szostak,M., Wanner,G. and Lubitz,W. (1990a) PhiX174 Protein E-Mediated Lysis of Escherichia coli. Biochimie 72: 191-200.

Witte,A., Wanner,G., Blasi,U., Halfmann,G., Szostak,M. and Lubitz,W. (1990b) Endogenous Transmembrane Tunnel Formation Mediated by ØX174 Lysis Protein E. J. Bacteriol. 172: 4109-4114.

Wretlind,B. and Pavlovskis,O.R. (1984) Genetic Mapping and Characterization of Pseudomonas aeruginosa Mutants Defective in the Formation of Extracellular Proteins. J. Bacteriol. 158: 801-808.

Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors. Gene 33: 103-119.

Zeitoun,F.M. and Wilson,E.E. (1968) Serological Comparisons of Erwinia nigrifluens with Certain Other Erwinia Species. Phytopathology 58: 1381-1385.

Zieg,J. and Simon,M. (1980) Analysis of the Nucleotide Sequence of an Invertible Controlling Element. Proc. Natl. Acad. Sci. USA 77: 4169-4200.

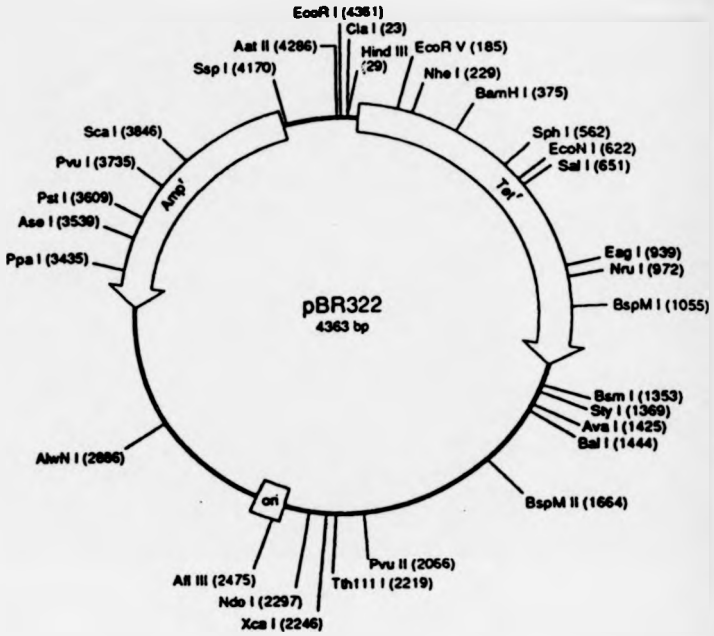
Zink,R.T. and Chatterjee,A.K. (1985) Cloning and Expression in Escherichia coli of Pectinase Genes of Erwinia carotovora Subsp. carotovora. Appl. Environ. Microbiol. 49: 714-717.

Zink,R.T., Kemble,R.J. and Chatterjee,A.K. (1984) Transposon Tn5 Mutagenesis in Erwinia carotovora subsp. carotovora and Erwinia carotovora subsp. atroseptica. J. Bacteriol. 157: 809-814.

Zubay,G. (1973) In Vitro Synthesis of Protein in Microbial Systems. Ann. Rev. Genet. 7: 267-287.

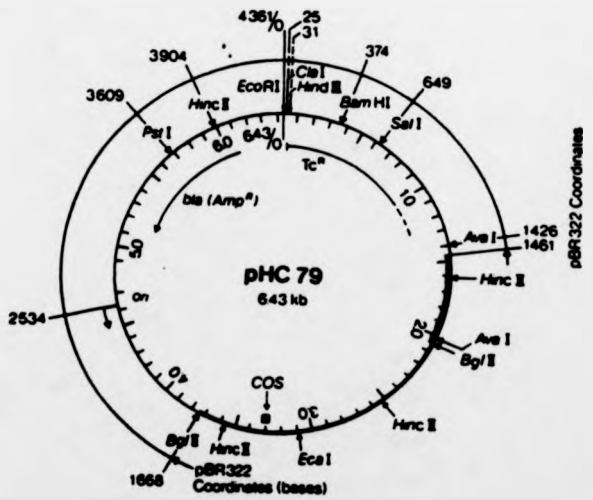
APPENDIX I

Restriction map of pBR322



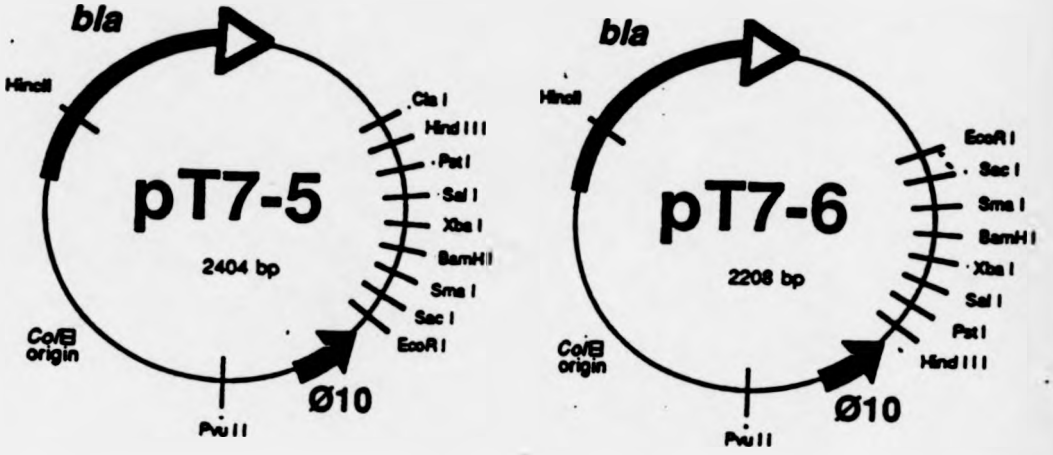
Bolivar et al. (1987)

Restriction map of pHC79

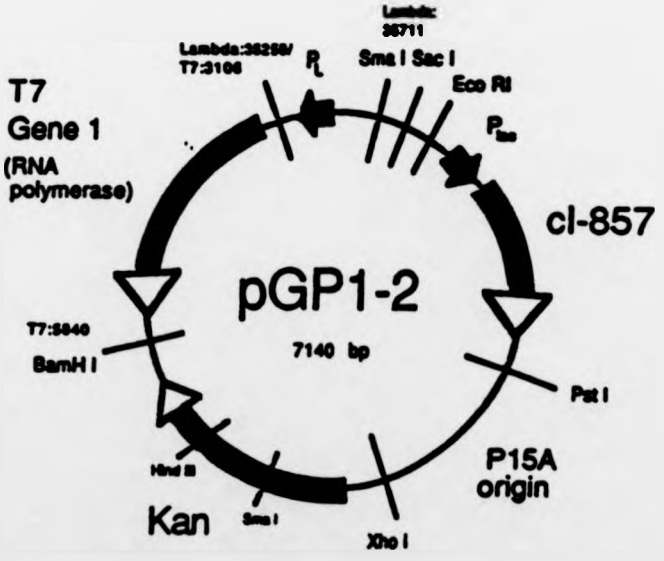


Hohn and Collins (1980)

T7 RNA polymerase/promoter vectors



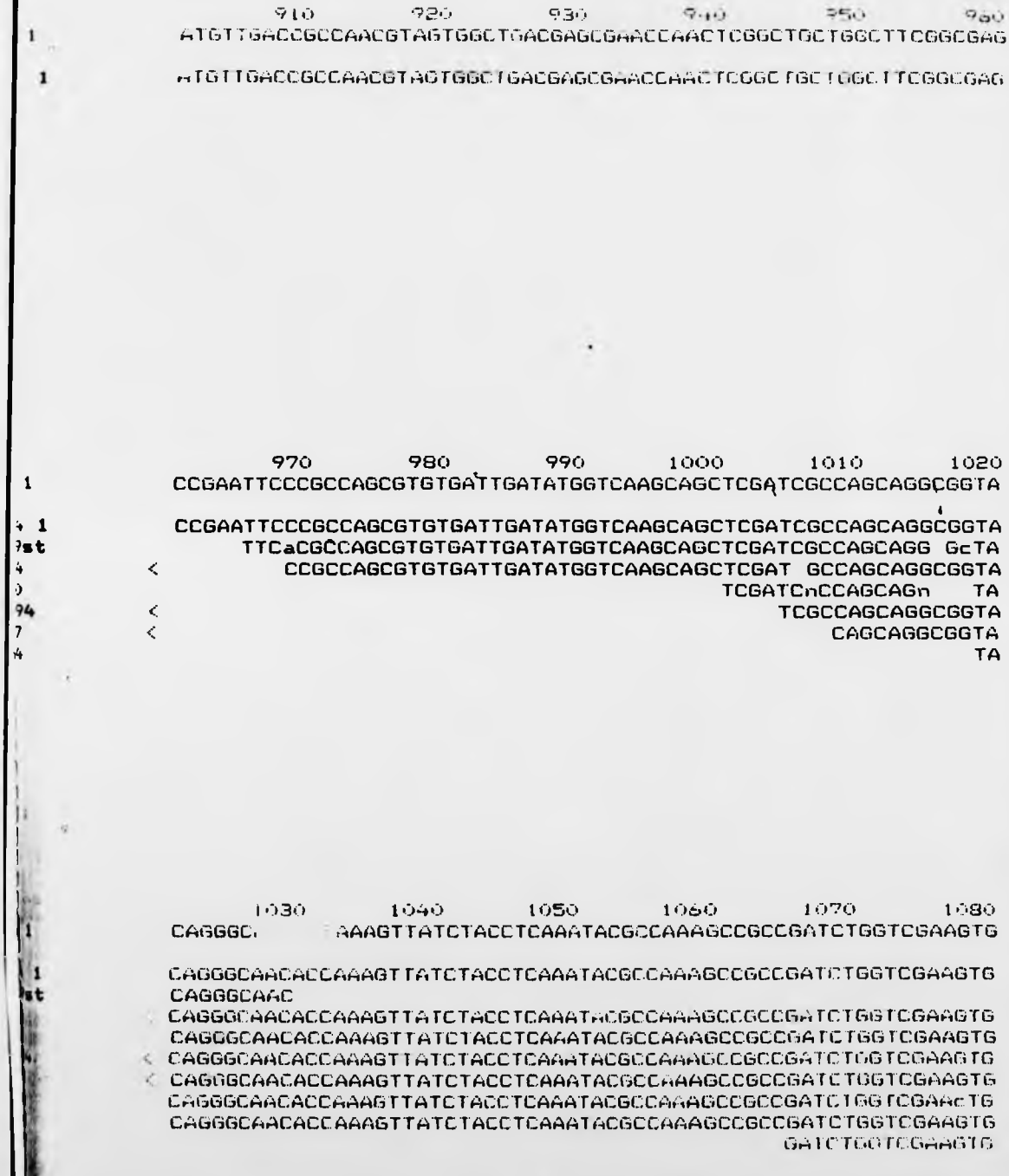
T7 RNA polymerase producing plasmid



Tabor and Richardson (1985)

APPENDIX II

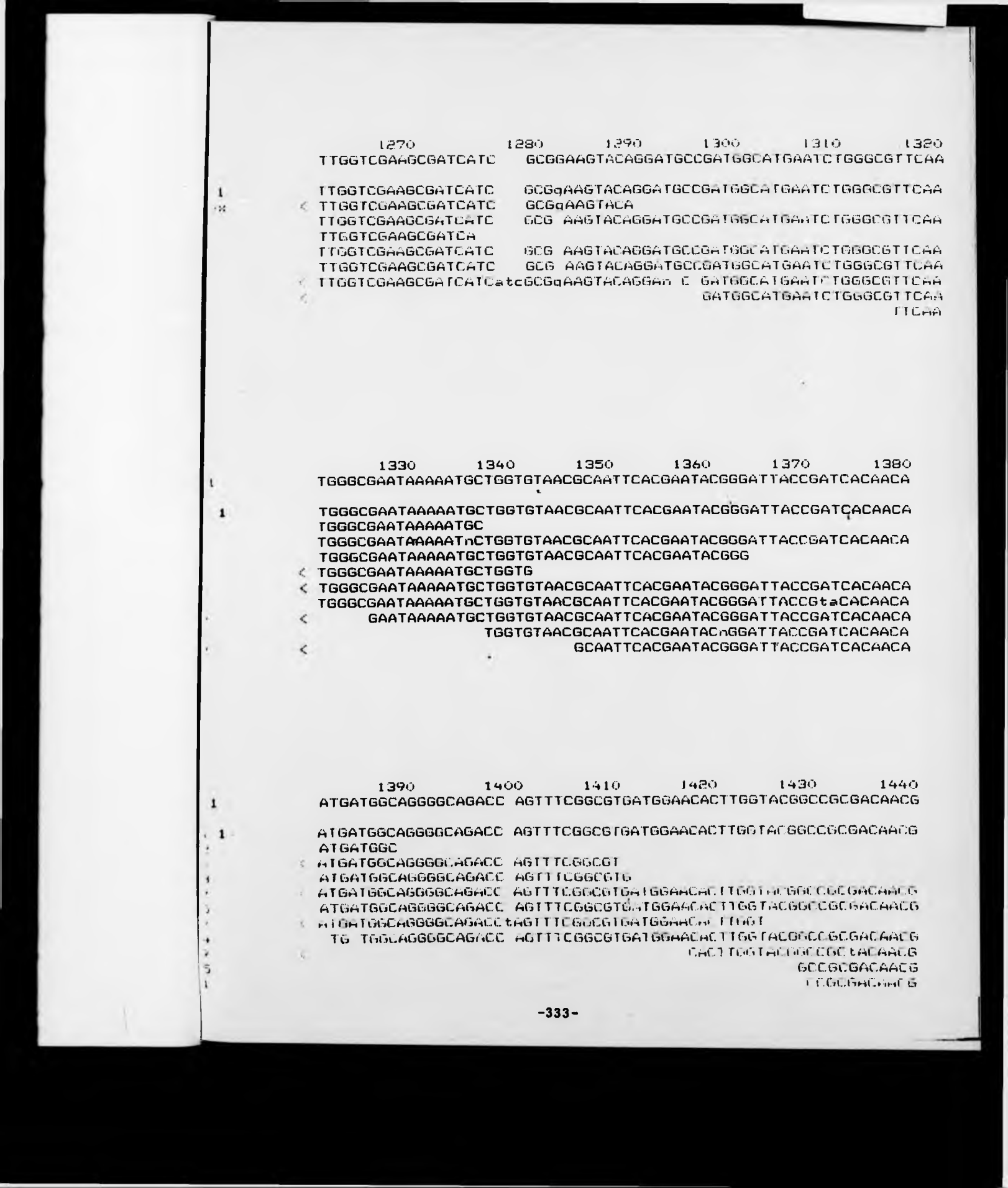
Appendix II shows the result of a "shotgun merge" (Queen and Korn, 1984) using the gel readings obtained from sequencing individual M13 clones. M13 clones containing DNA which was prone to "compression" (G:C secondary structure formation during electrophoresis) were re-sequenced substituting dGTP for dITP. Such regions have been underlined in this compilation.


910 920 930 940 950 960
ATGTTGACCGCCAACGTAGTGGCTGACGAGCGAACCAACTCGGCTGCTGGCTTCGGCGAG
970 980 990 1000 1010 1020
CCGAATTCCCGCCAGCGTGTGATTGATATGGTCAAGCAGCTCGATCGCCAGCAGGCGGTA
1030 1040 1050 1060 1070 1080
CAGGGCAACACCAAAGTTATCTACCTCAAATACGCCAAAGCCGCCGATCTGGTCGAAGTG

```
        1090        1100        1110        1120        1130        1140
CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC

1       CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
<       CTCACCGGTGTCGG CGACAGTATCCAAACC
        CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
<       CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
<       CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAA
        CTCACCGGTGTCGGaCGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
        CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
        CTCACCGGTGTCGG CGACAGTATCCAAACCGATCAGCAAAATGCGCTGCCTGCACTGCGC
:       <                                 AaCGATCAGCAAA TGCGCTGCCTGCACTGCGC
```

```
        1150        1160        1170        1180        1190        1200
AAAGACATTTCGATTAAGGCACACGAACAAACCAACTCGCTGATTGTGAATGCCGCACCG

AAAGACATTTCGATTAAGGCACACGAACAAACCAACTCGCTGATTGTGAATGCCGCACCG
AAAGACATTTCGATTAAGGCACA
< AAAGACATTTCGATT
AAAGACATTTCGATTAA
AAAGACATTTCGATTAAGGCACAC AACAAACCAACTC CTGATTGTGAATGCCGCACCG
AAAGACATTTCGATTAAGGCACACGAACAAACCAACTCGCTGA
< AA GACATTTCGATTAAGGCACACGAACAAACCAACTCGCTGATTGTGAATGCCGCACCG
AAAGACATTTCGATTAAGGCACACGAACAAACCAACTCGCTGATTGTGAATGCCGCACCG
            GATTAAGGCACACGAACAAACCAACT GCTGATTGTGAATGCCGCACCG
```

```
        1210        1220        1230        1240        1250        1260
GA CA TCATGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAGGTG

GA CA TCATGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAGgTG
GA CA TCATGCGC
< GA CA TCATGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAGqTG
GAaCAaTCATGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAG TG
GA CA TCATGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAGtqG
          TGCGCGATCTGGAACAGGTGATTGCGCAGTTGGATATCCGTCGTCCGCAGqTG
                TGGAACAG TGATTGCGCAGTTGGATATCCGTCGTCCGCAG TG
                                                CGTCCGCAG TG
```



-333-

```
      1450        1460        1470        1480        1490        1500
GCACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCTG

GCACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCTG
< GCA
GCACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCT
GCACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCTG
GCACTTGGCGGTTTCAA GcCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCTG
cgACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGCATGCTG
n ACTTGGCGGTTTCAACGGCATTGCTGCnG TTTCTATCAGGGTAACTGGGGCATGCTG
GCACTTGGCGGTTTCAACGGCATTGCTGCCGGTTTCTATCAGGGTAACTGGGGC TGCTG
<                                                   GCATGCTG
<                                                   GCATGCTG
```

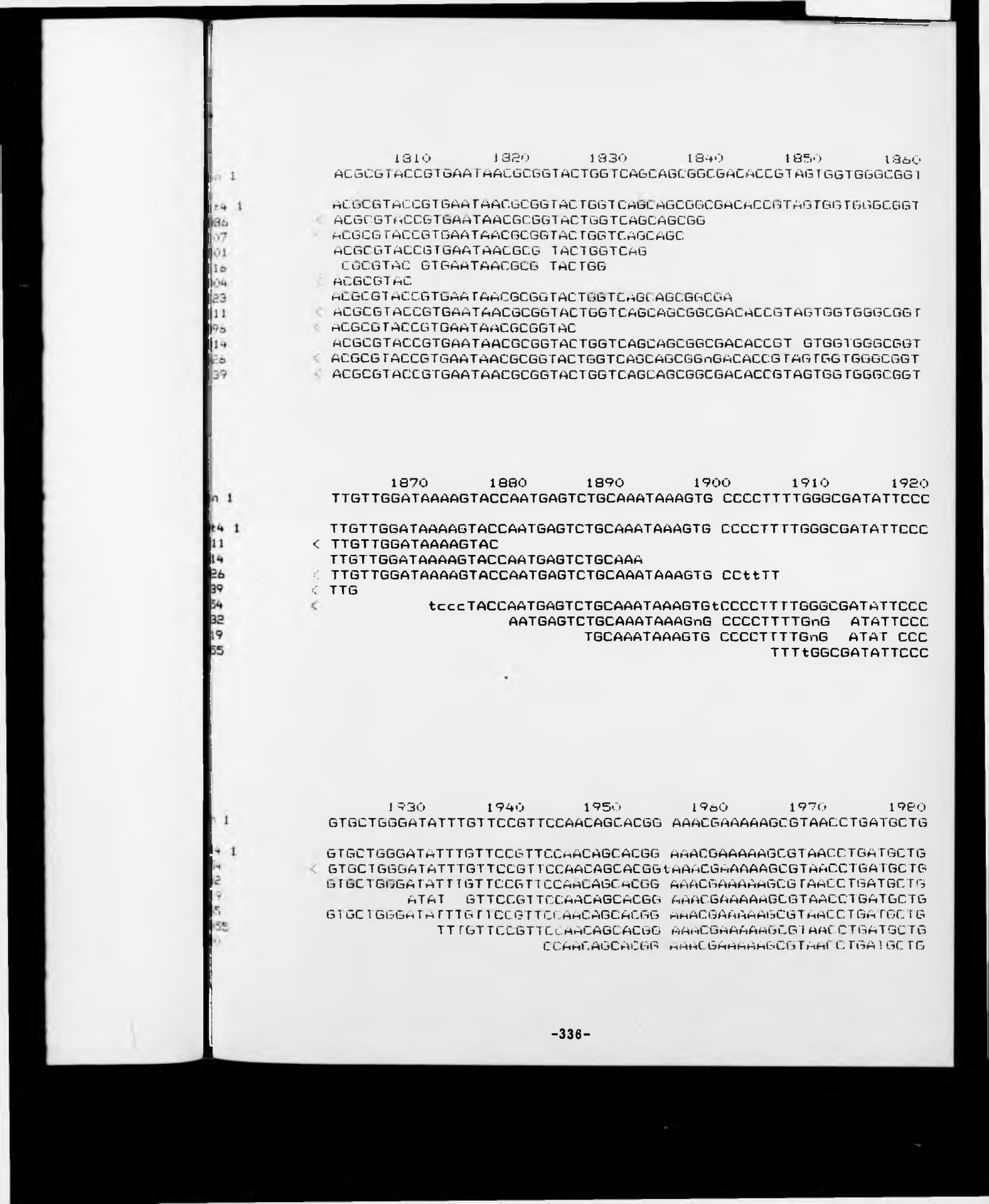
```
      1510        1520        1530        1540        1550        1560
ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTGGCAAC GCCCAGTA TTGTGACG

ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTGGCAAC GCCCAGTA TTGTGACG
gTG  GGaAtTcg
A
< ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTG CAAn GCCCAGTA TTGTGACG
  ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTGGCAAC GCC AGTA TTGTGACG
  ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTGGCAAC GCC AGTA TTGTGACG
< ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTGGCAACtGCCCAGTA TTGTGACG
< ATGACGGCACTG TCCAGCAACAGTAAAAACGATATTCTG CAAC GCCCAGTA TTGTGACG
< ATGACG CACTG TCCAGCAACAGTAAAAACGATATTCTG CAAC GCCCAGTA TTGTGACG
               CCAGCAAC GTAAAAACGATATTCTGGCAACn C  AGTAaTTGTGACG
```

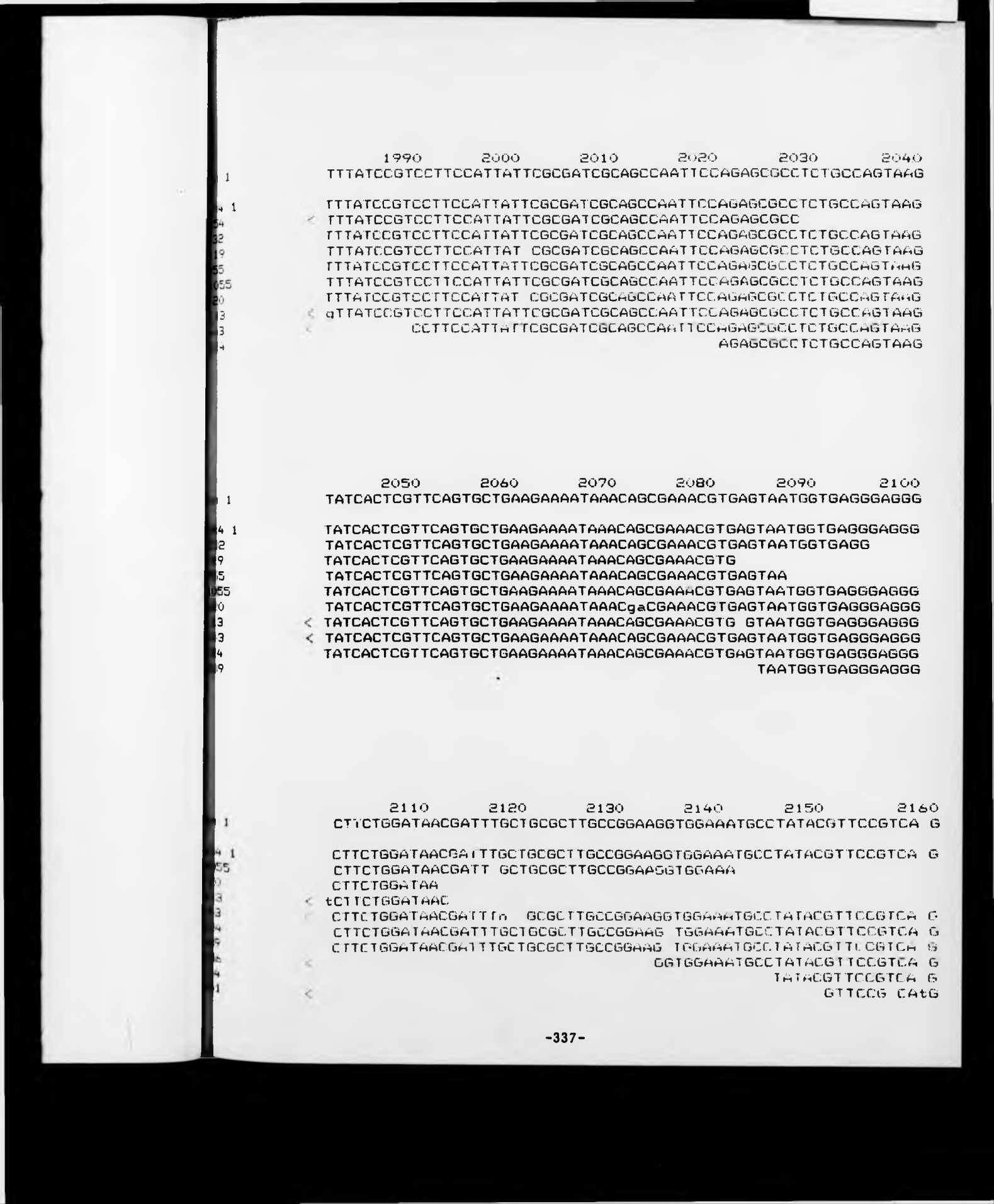
```
      1570        1580        1590        1600        1610        1620
CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG

CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG
< CTGGACAATATGGAGGCAA
  CTGGACAATATG AGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG
  CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCC
  CTGGACAATATGGAGGCAACGTTT
  CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG
< CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG
  CTGGACAATATGGAGGCAACGTTTAACGTCGGTCAGGAAGTGCCAGTATTGGCCGGTTCG
                                                    CG
```

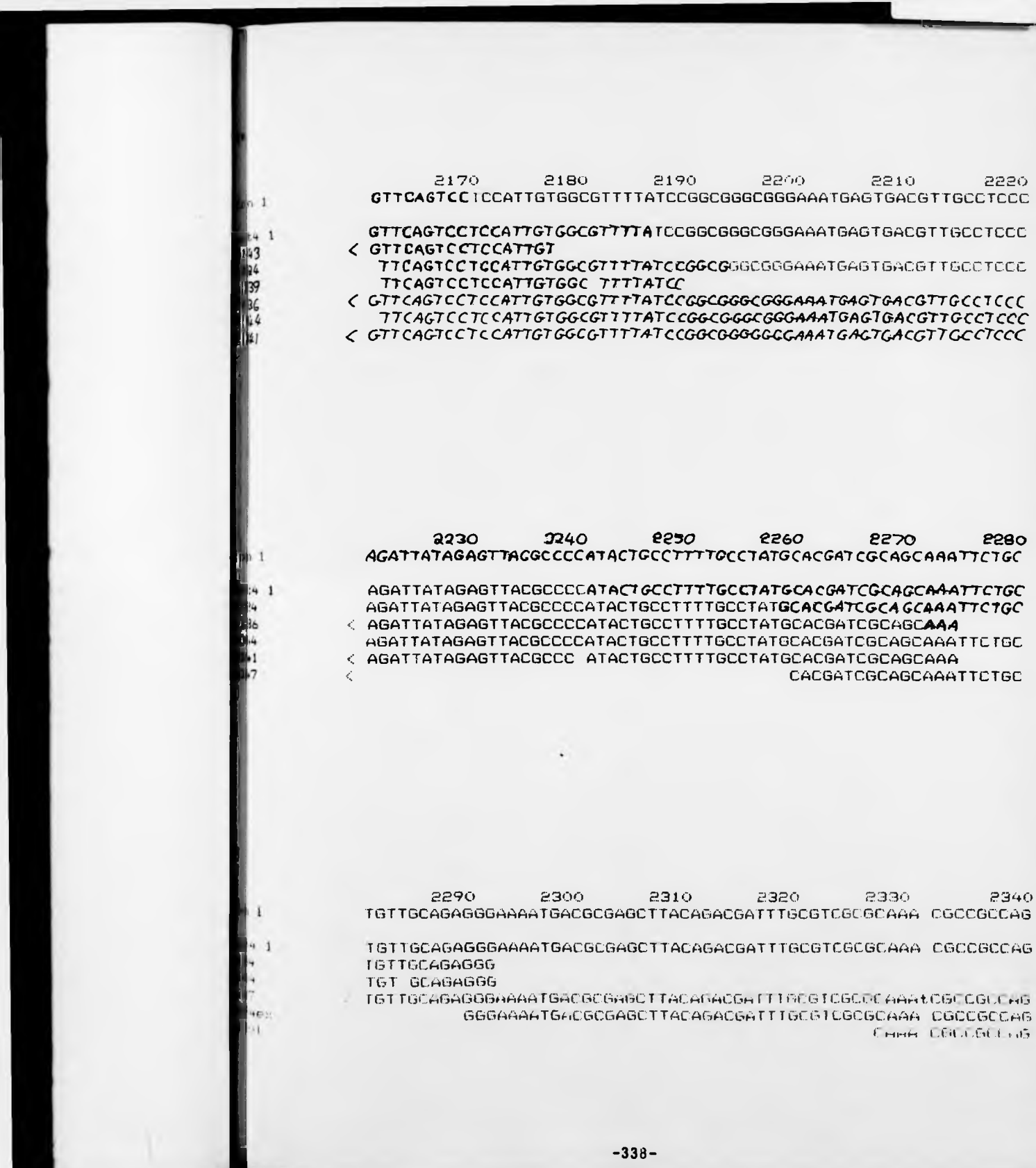
```
        1630        1640        1650        1660        1670        1680
CAGACGACATCCGGTGATAACGTTTTCCAAACCGTGGAACGTA AAACGGTCGGTATCAAG

1       CAGnCGnCATCCGGTGATAACGTTTTCCAAACCGTGGAACGTn AAACGGTCGGTnTCAAG
        CAGACGACATCC
        CAGnCGACATCCGGTGnTAACGTTTTCCAnAACCGTGGAACGTA AAACGGTCGGTnTCAA
        CAGACGACATCCGGTGATAACGTTTTCCAAACCGTGGAACGTA AAACGGTCGGTATCAAG
        CAGACGACATCCGGTGATAACGTTTTCCAAAC GTGGAACGTA AAACGGTCGGTATCAAG
        CAGnCGACATCCGGTGATAACGTTTTCCAAACCGTGGAACGTA AAACGGTCGGTATCAAG
            CGACATCCGGTGATAACGTTTTCCAAACCGTGGAACGTA AAACGGTCGGTATCAAG
                ATCCGGTGATAACGTTTTCCAAACCGTGGAACGTA AAACGGTCGGTATCAAG
                        AACGTTTTCCAAACCGTGGAACGTA AAACGGTCGGTATCAAG
                                ACCGTGGAACGTA AAACGnTnnGTATCAAG
                                   GTGGAACGTA AAACGGTCGGTATCAAG
                                   GTGGAACGTAgAAACGGTCGGTATCAAG
                                              CGGTCGGTATCAAG
                                                 CGGTATCAAG
```

```
        1690        1700        1710        1720        1730        1740
CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA

1       CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGT
        CTGAAGGTGAAACCCaAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCCCAAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
        CTGAAGGTGAAACCC AAATCAATGAAGGTGATTCCGTGTTGCTGGAGATCGAGCAGGAA
                                   GTGTTGCTGGAGATCGAGCAGGAA
                                   9
```

```
        1750        1760        1770        1780        1790        1800
GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT

1       GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
        GTCTCCaaCGTGGCAGACGCGGCCTCC
        GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
        GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGG
        GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGAnGTTCAAT
        GTCTCCAGCnTGGCAGACGCnGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
        GT TCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACC
        GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
        GnCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGnGACGTTCAAT
        GTCTCCAGCGTGGCAGAC CGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
        GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
        GTCTCCAGCGTGGCAGACGCGGCCTCCAGCAGCAGCACCAACCTCt AGCGAnGTTCAAT
            CCAGCGTGGCAGACn GGCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
                            GCCTCCAGCAGCAGCACCAACCTCGGAGCGACGTTCAAT
                                                    GTTCAAT
```


1810 1820 1830 1840 1850 1860
ACGCGTACCGTGAATAACGCGGTACTGGTCAGCAGCGGCGACACCGTAGTGGTGGGCGGT
1870 1880 1890 1900 1910 1920
TTGTTGGATAAAAGTACCAATGAGTCTGCAAATAAAGTG CCCCTTTTGGGCGATATTCCC
1930 1940 1950 1960 1970 1980
GTGCTGGGATATTTGTTCCGTTCCAACAGCACGG AAACGAAAAAGCGTAACCTGATGCTG

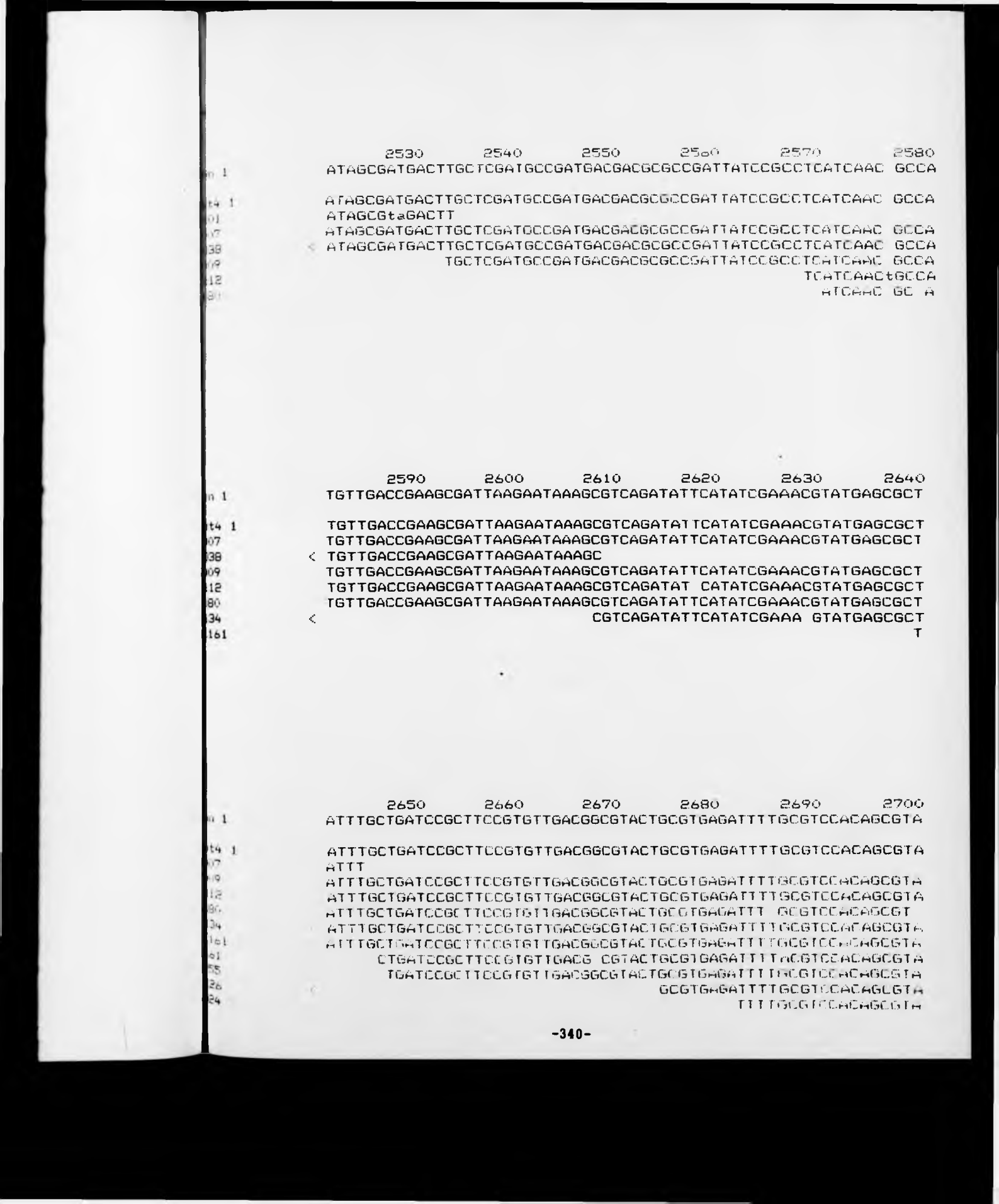

1990 2000 2010 2020 2030 2040
TTTATCCGTCCTTCCATTATTCGCGATCGCAGCCAATTCCAGAGCGCCTCTGCCAGTAAG
TTTATCCGTCCTTCCATTATTCGCGATCGCAGCCAATTCCAGAGCGCC
TTTATCCGTCCTTCCATTATTCGCGATCGCAGCCAATTCCAGAGCGCCTCTGCCAGTAAG
TTTATCCGTCCTTCCATTAT CGCGATCGCAGCCAATTCCAGAGCGCCTCTGCCAGTAAG
qTTATCCGTCCTTCCATTATTCGCGATCGCAGCCAATTCCAGAGCGCCTCTGCCAGTAAG
CCTTCCATTATTCGCGATCGCAGCCAATTCCAGAGCGCCTCTGCCAGTAAG
AGAGCGCCTCTGCCAGTAAG
2050 2060 2070 2080 2090 2100
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGGGAGGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGGGAGGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAA
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGGGAGGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACgaCGAAACGTGAGTAATGGTGAGGGAGGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTG GTAATGGTGAGGGAGGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGGGAGGG
TATCACTCGTTCAGTGCTGAAGAAAATAAACAGCGAAACGTGAGTAATGGTGAGGGAGGG
TAATGGTGAGGGAGGG
2110 2120 2130 2140 2150 2160
CTTCTGGATAACGATTTGCTGCGCTTGCCGGAAGGTGGAAATGCCTATACGTTCCGTCA G
CTTCTGGATAACGATTTGCTGCGCTTGCCGGAAGGTGGAAATGCCTATACGTTCCGTCA G
CTTCTGGATAACGATT GCTGCGCTTGCCGGAAGGTGGAAA
CTTCTGGATAA
tCTTCTGGATAAC
CTTCTGGATAACGATTTn GCGCTTGCCGGAAGGTGGAAATGCCTATACGTTCCGTCA G
CTTCTGGATAACGATTTGCTGCGCTTGCCGGAAG TGGAAATGCCTATACGTTCCGTCA G
CTTCTGGATAACGATTTGCTGCGCTTGCCGGAAG TGGAAATGCCTATACGTTCCGTCA G
GGTGGAAATGCCTATACGTTCCGTCA G
TATACGTTCCGTCA G
GTTCCG CAtG

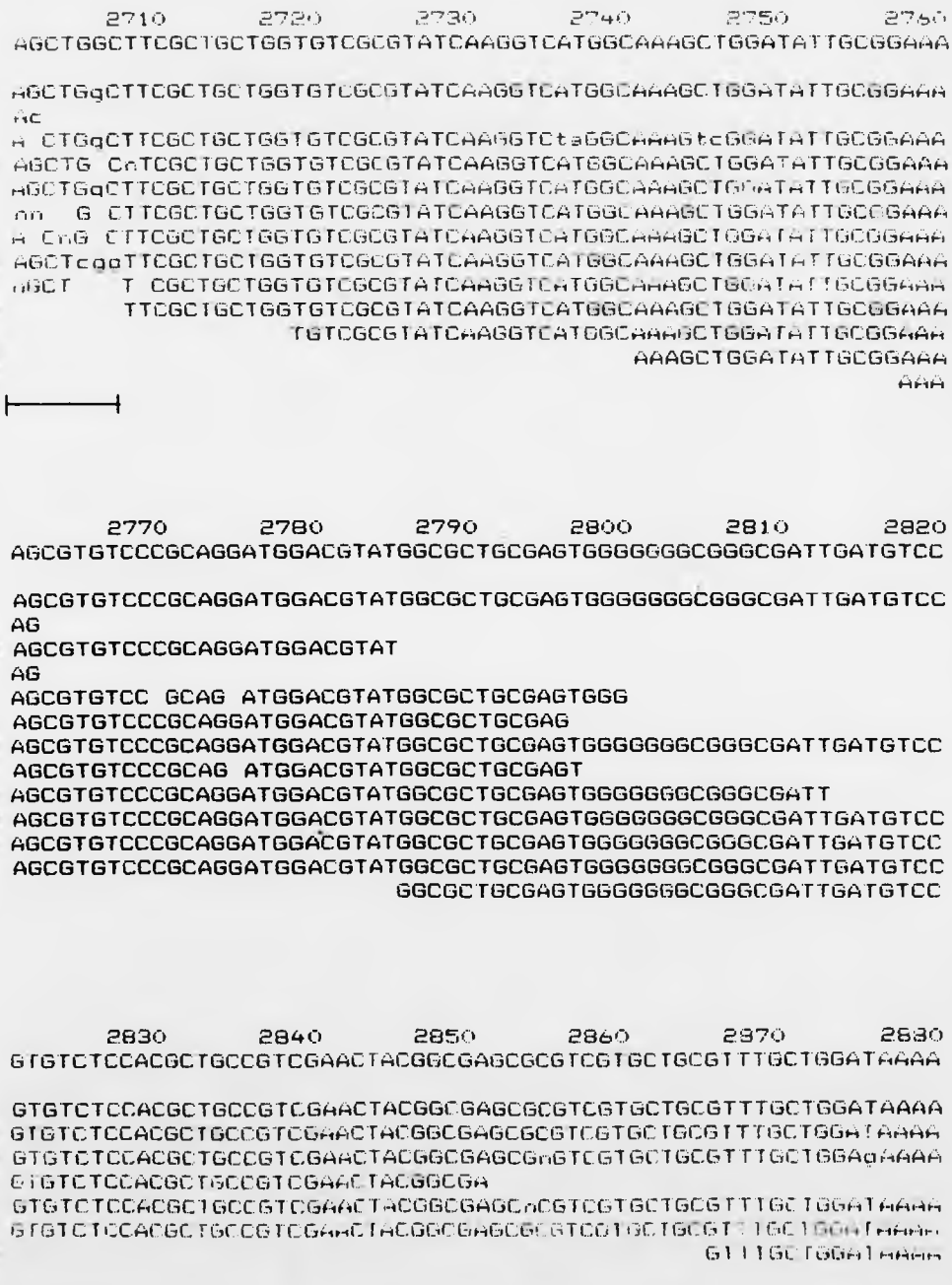

2170 2180 2190 2200 2210 2220
GTTCAGTCC TCCATTGTGGCGTTTTATCCGGCGGGCGGGAAATGAGTGACGTTGCCTCCC
GTTCAGTCCTCCATTGTGGCGTTTTATCCGGCGGGCGGGAAATGAGTGACGTTGCCTCCC
< GTTCAGTCCTCCATTGT
TTCAGTCCTCCATTGTGGCGTTTTATCCGGCGGGCGGGAAATGAGTGACGTTGCCTCCC
TTCAGTCCTCCATTGTGGC TTTTATCC
< GTTCAGTCCTCCATTGTGGCGTTTTATCCGGCGGGCGGGAAATGAGTGACGTTGCCTCCC
TTCAGTCCTCCATTGTGGCGTTTTATCCGGCGGGCGGGAAATGAGTGACGTTGCCTCCC
< GTTCAGTCCTCCATTGTGGCGTTTTATCCGGCGGGGGCGAAATGAGTGACGTTGCCTCCC
2230 2240 2250 2260 2270 2280
AGATTATAGAGTTACGCCCCATACTGCCTTTTGCCTATGCACGATCGCAGCAAATTCTGC
AGATTATAGAGTTACGCCCCATACTGCCTTTTGCCTATGCACGATCGCAGCAAATTCTGC
AGATTATAGAGTTACGCCCCATACTGCCTTTTGCCTATGCACGATCGCAGCAAATTCTGC
< AGATTATAGAGTTACGCCCCATACTGCCTTTTGCCTATGCACGATCGCAGCAAA
AGATTATAGAGTTACGCCCCATACTGCCTTTTGCCTATGCACGATCGCAGCAAATTCTGC
< AGATTATAGAGTTACGCCC ATACTGCCTTTTGCCTATGCACGATCGCAGCAAA
< CACGATCGCAGCAAATTCTGC
2290 2300 2310 2320 2330 2340
TGTTGCAGAGGGAAAATGACGCGAGCTTACAGACGATTTGCGTCGCGCAAA CGCCGCCAG
TGTTGCAGAGGGAAAATGACGCGAGCTTACAGACGATTTGCGTCGCGCAAA CGCCGCCAG
TGTTGCAGAGGG
TGT GCAGAGGG
TGTTGCAGAGGGAAAATGACGCGAGCTTACAGACGATTTGCGTCGCGCAAA CGCCGCCAG
GGGAAAATGACGCGAGCTTACAGACGATTTGCGTCGCGCAAA CGCCGCCAG
CAAA CGCCGCCAG

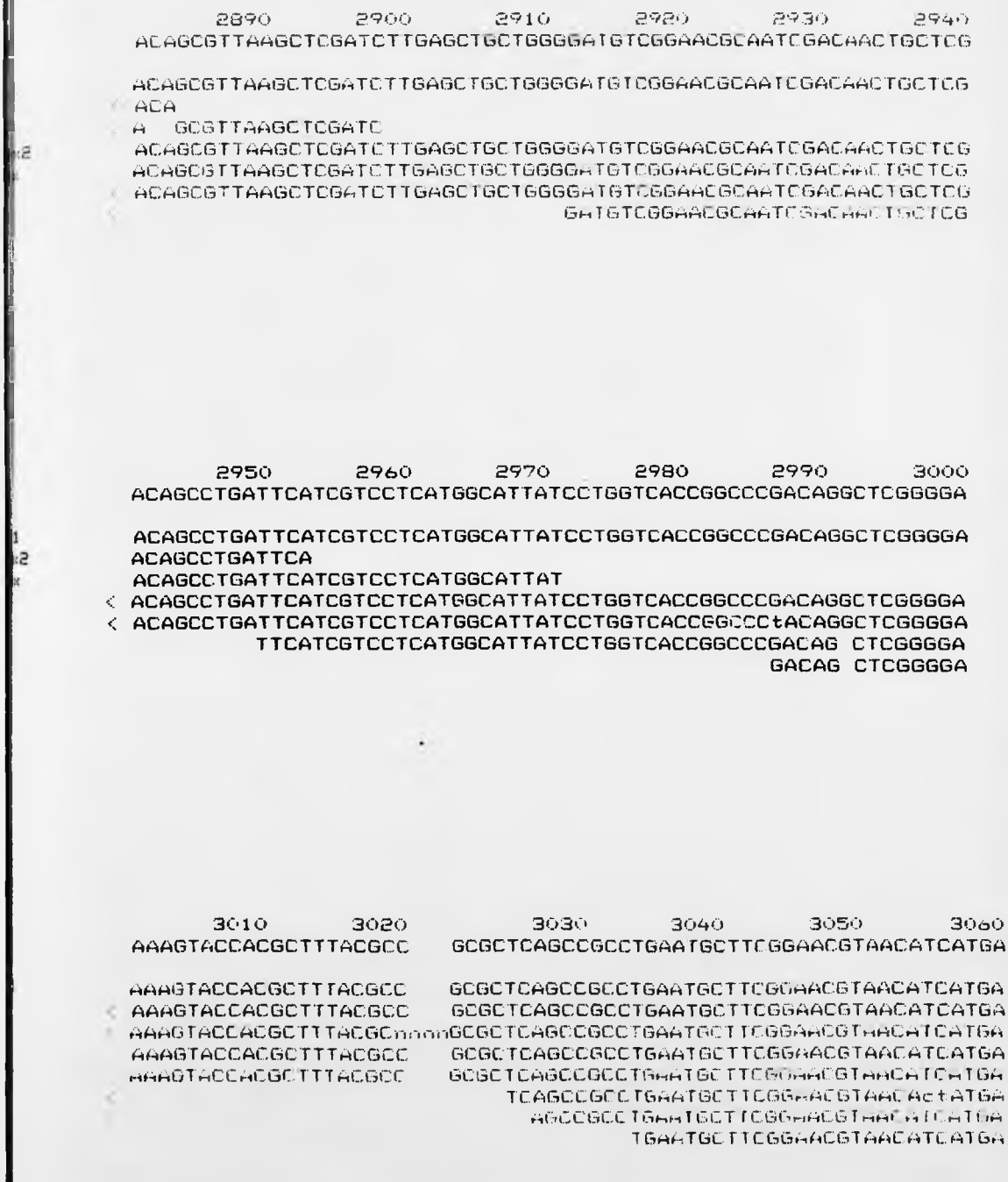
```
        2350        2360        2370        2380        2390        2400
CCGCTTTGCTGGAAGCGC GTCGGATTGCAGGCTGTTCGCTCAGGATTGAGCGCGTTACGG

CCGCTTTGCTGGAAGCGC GTCGGATTGCAGGCTGTTCGCTCAGGATTGAGCGCGTTACGG
C GCTTTGCTGGAAGCGCtGTCGGATTGCAGGCTGTTCGCTCAGGATTGAGCGnGTTACGG
CCGCTTT
CCGCTTTGCTGGAAGCGC GTCGGATTGCAGGCTGTTCGCTCAGGATTGAGCGCGTTACGG
CCGCTTTGCTGGAAGCGn GTCGGATTGCAGGCTGTTCGCTCAGGATTGAGCGnGTTACGG
                  tC GTCGGATTGCAGGCTGTTCGCTCAGGATTGAGC CGTTACGG
```

```
        2410        2420        2430.       2440        2450        2460
ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGACTCGGAAGAAGCGCGCC

ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGACTCGGAAGAAGCGCGCC
ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGACTCGGAAGAAGCGnGCC
ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGAC CGGAAGAAGCGCGCC
ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGACTCGGAAGAAGCGCGCC
ATGAAGAATTTGAGCGGCAATTAGTCATTAGCTATCAGCGCGACTCGGAAGAAGCGCGCC
                                          ACTCGGAAGAAGCGCGCC
```

```
        2470        2480        2490        2500        2510        2520
GTATGATGGAGGACATTGGTAATGAGATGGACTTCTATACGCTGGTGGAAGAACTACCAG

GTATGATGGAGGACATTGGTAATGAGATGGACTTCTATACGCTGGTGGAAGAACTACCAG
GTATGATGGA
GTA
GTATGATGGA
GTATGATGGAGGACATTGGTAATGAGATGGACTTCTATACGCTGGTGGAAGAACTACCAG
GTATGATGGAGGACATTGGTAATGAGATGGACTTCTATACGCTGGTGGAAGAACTACCAG
        GGAGGACATTGGTAATGAGATGGACTTCTATACGCTGGTGGAAGAACTACCAG
```


2530 2540 2550 2560 2570 2580
ATAGCGATGACTTGCTCGATGCCGATGACGACGCGCCGATTATCCGCCTCATCAAC GCCA
ATAGCGATGACTTGCTCGATGCCGATGACGACGCGCCGATTATCCGCCTCATCAAC GCCA
ATAGCGtaGACTT
ATAGCGATGACTTGCTCGATGCCGATGACGACGCGCCGATTATCCGCCTCATCAAC GCCA
ATAGCGATGACTTGCTCGATGCCGATGACGACGCGCCGATTATCCGCCTCATCAAC GCCA
TGCTCGATGCCGATGACGACGCGCCGATTATCCGCCTCATCAAC GCCA
TCATCAACtGCCA
ATCAAC GC A
2590 2600 2610 2620 2630 2640
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATATTCATATCGAAACGTATGAGCGCT
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATATTCATATCGAAACGTATGAGCGCT
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATATTCATATCGAAACGTATGAGCGCT
TGTTGACCGAAGCGATTAAGAATAAAGC
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATATTCATATCGAAACGTATGAGCGCT
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATAT CATATCGAAACGTATGAGCGCT
TGTTGACCGAAGCGATTAAGAATAAAGCGTCAGATATTCATATCGAAACGTATGAGCGCT
CGTCAGATATTCATATCGAAA GTATGAGCGCT
T
2650 2660 2670 2680 2690 2700
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
ATTT
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTT GCGTCCACAGCGT
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
ATTTGCTGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
CTGATCCGCTTCCGTGTTGACG CGTACTGCGTGAGATTTTGCGTCCACAGCGTA
TGATCCGCTTCCGTGTTGACGGCGTACTGCGTGAGATTTTGCGTCCACAGCGTA
GCGTGAGATTTTGCGTCCACAGCGTA
TTTTGCGTCCACAGCGTA


2710 2720 2730 2740 2750 2760
AGCTGGCTTCGCTGCTGGTGTCGCGTATCAAGGTCATGGCAAAGCTGGATATTGCGGAAA
AGCTGGCTTCGCTGCTGGTGTCGCGTATCAAGGTCATGGCAAAGCTGGATATTGCGGAAA
2770 2780 2790 2800 2810 2820
AGCGTGTCCCGCAGGATGGACGTATGGCGCTGCGAGTGGGGGGGCGGGCGATTGATGTCC
AGCGTGTCCCGCAGGATGGACGTATGGCGCTGCGAGTGGGGGGGCGGGCGATTGATGTCC
2830 2840 2850 2860 2870 2880
GTGTCTCCACGCTGCCGTCGAACTACGGCGAGCGCGTCGTGCTGCGTTTGCTGGATAAAA
GTGTCTCCACGCTGCCGTCGAACTACGGCGAGCGCGTCGTGCTGCGTTTGCTGGATAAAA


2890 2900 2910 2920 2930 2940
ACAGCGTTAAGCTCGATCTTGAGCTGCTGGGGATGTCGGAACGCAATCGACAACTGCTCG
ACAGCGTTAAGCTCGATCTTGAGCTGCTGGGGATGTCGGAACGCAATCGACAACTGCTCG
ACA
A GCGTTAAGCTCGATC
ACAGCGTTAAGCTCGATCTTGAGCTGCTGGGGATGTCGGAACGCAATCGACAACTGCTCG
ACAGCGTTAAGCTCGATCTTGAGCTGCTGGGGATGTCGGAACGCAATCGACAACTGCTCG
ACAGCGTTAAGCTCGATCTTGAGCTGCTGGGGATGTCGGAACGCAATCGACAACTGCTCG
GATGTCGGAACGCAATCGACAACTGCTCG
2950 2960 2970 2980 2990 3000
ACAGCCTGATTCATCGTCCTCATGGCATTATCCTGGTCACCGGCCCGACAGGCTCGGGGA
ACAGCCTGATTCATCGTCCTCATGGCATTATCCTGGTCACCGGCCCGACAGGCTCGGGGA
ACAGCCTGATTCA
ACAGCCTGATTCATCGTCCTCATGGCATTAT
ACAGCCTGATTCATCGTCCTCATGGCATTATCCTGGTCACCGGCCCGACAGGCTCGGGGA
ACAGCCTGATTCATCGTCCTCATGGCATTATCCTGGTCACCGGCCCtACAGGCTCGGGGA
TTCATCGTCCTCATGGCATTATCCTGGTCACCGGCCCGACAG CTCGGGGA
GACAG CTCGGGGA
3010 3020 3030 3040 3050 3060
AAAGTACCACGCTTTACGCC GCGCTCAGCCGCCTGAATGCTTCGGAACGTAACATCATGA
AAAGTACCACGCTTTACGCC GCGCTCAGCCGCCTGAATGCTTCGGAACGTAACATCATGA
AAAGTACCACGCTTTACGCC GCGCTCAGCCGCCTGAATGCTTCGGAACGTAACATCATGA
AAAGTACCACGCTTTACGCC GCGCTCAGCCGCCTGAATGCTTCGGAACGTAACATCATGA
AAAGTACCACGCTTTACGCC GCGCTCAGCCGCCTGAATGCTTCGGAACGTAACATCATGA
TCAGCCGCCTGAATGCTTCGGAACGTAACActATGA
AGCCGCCTGAATGCTTCGGAACGTAACATCATGA
TGAATGCTTCGGAACGTAACATCATGA


3070 3080 3090 3100 3110 3120
CGGTGGAAGATCCCATCGAGTATGAACTGGAGGGTATCGGGCAAACGCAGGTCAACACCA
CGGTGGAAGATCCCATCGAGTATGAACTGGA
CGGTGGAAGATCCCATCGAGTATGAACTGGA
CGGTGGAAGATCCCATCGAGTATGAACTG AGGGTATCGGGCAAACGCAGGTCAACACCA
CGGTGGAAGATCC ATCGAGTATGAACTGGAGGG ATCGGGCAAACGCAGGTCAACACCA
CGGTGGAAGATCCCATCGAGTATGAACTGGAGGGTATCGGGCAAACGCAGGTCAACACCA
CGGTGGAAGATCCCATCGAGTATGAACTGGAGGGTATCGGGCAAACGCAGGTCAACACCA
CGGTGGAAGATCCCATCGAGTATGAACTGGAGGGTATCGGGCAAACGCAGGTCAACACCA
3130 3140 3150 3160 3170 3180
AGGTCGATATGACGTTTGCCCGCGGGCTGCGTGCCATTCTGCGTCAGGACCCGGACGTCG
AGGTCGATATGACGTTTGCCCGCGGGCTGCGTGCCATTCTGCGTCAGGACCCGGACGTCG
AGGTC
AGGTCGATATGACGTTTcgCnGCGGGCTGCGTGCCATTCTGCGTCAGGACCCGGACGTCG
AGGTCGATATGACGTTT
AGGTCGATATGACGTTTGCCCGCGGGCTGCGTGCCATTCTGCGTCAGGACCCGGACGTCG
3190 3200 3210 3220 3230 3240
TGCTGGTGGGGGAAATTCGTGATGGTGAAACGGCGCAGATTGCCGTGCAGGCCTCGTTGA
TGCTGGTGGGGGAAATTCGTGATGGTGAAACGGCGCAGATTGCCGTGCAGGCCTCGTTGA
TGCTGGTGGGGGAAATTCGTGATGGTGAAAC
TGCTGGTGGGGGAAATTCGTGATGGTGAAACGGCGCAGATTGCCGT
GGGGGAAnTTCGTGATGGTGAAACGGCGCAGATTGCCGTGCAGGCCTCGTTGA
GTGATGGTGAAACGGCGCAGATTGCCGTGCAGnC TCGTTGA
GAAACGGCGCAGAT GCCGTGCAnGCCTCGTTGA
AACGGCGCAGATTGCCGTGCAGnC TCGTTGA
TCGTTGA

```
       3250        3260        3270        3280        3290        3300
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGCGCGCTGTCCCGTT

CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGCGCGCTGTCCCGTT
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGCGCGCTGTCCCGTT
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGCGCGCTGTCCCGTT
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGCGCGCTGTCCCGTT
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGC     TGTCCCGTT
CCGGTCACCTCGTGTTATCCACACTGCATACCAATAGTGCGCTGGGC     TGTCCCGTT
```

```
       3310        3320        3330        3340        3350        336
TG CAGGATATGGGCGTTGAGCCTTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCG

TG CAGGATATGGGCGTTGAGCCTTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCG
TG CAGGATATGGGCGTTGAGCCTTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCG
Ttg
TG CAGGATATGGGCGTTGAGCCTTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCG
Tn CAGGATATGGGCGTTGAGC TTTCCTGCTGTCAACCTCTCTACTGGG
TG CAGGATATGGGCGTTGAGC TTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCG
                       TTTCCTGCTGTCAACCTCTCTACTGGGCGTACTCGCG
                       TTTggTGCTGTCAACCTCTCTACTGGGCGTACTCGCn
```

```
      3370        3380        3390        3400        3410        3420
AGCGTCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC

AGCGTCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
AGCGTCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
AGCG1C GGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
An  TCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGtaCCGGT C
AGCGTCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
An  TCTGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
      TGGTCAGGACGCTGTGTTCTGACTGTAGCCAGCCGCAACCTGTCGATCCGGTTC
                                                GATCCGGTTC
```

```
        3430        3440        3450        3460        3470        3480
1       AGGCTGAACAGATGGGGATCGCGCCCGGTACG CTACTGCATAACCCCGTTGGCTGTCCGC

 1      AGGCTGAACAGATGGGGATCGCGCCCGGTACG CTACTGCATAACCCCGTTGGCTGTCCGC
me      AGG
3       H
        AGGCTGAACAGATGGGGtaCGCGCCC
02      AGGCTGAACAGATGGGGtaCGCGCnCGGTACGtCTAC GCATAACCCCGTTGG
        AGGCTGAACAGATGGGGATCGCGCCCGGTACG CTACTGCATAACCCCGTTGGCTGTCCGC
        AGGCTGAACAGATGGGGATCGCGCCCGGTACG CTACTGCATAACCCCGTTGGCTGTCCGC
        AGGCTGAACAGATGGGGATCGCGnCCGGTACG CTACTGCATAACCCCGTTGGCTGTCCGC
5       AGGCTGAACAGATGGGGATCGCGCCCGGTACG CTACTGCATAACCCCGTTGGCTGTCCGC
                                                TGCATAACCCCGTTGGCTGTCCGC
                                                      TTGGCTGTCCGC
ex                                                          GC


        3490        3500        3510        3520        3530        3540
1       AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCATGAACTGG TGCTGATTAATG

 1      AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCATGAACTGG TGCTGATTAATG
        AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCATGAACTGG TGCT
        AGTGTAGCTTTACCGGCTACCGGG  ACGTATCGGCATTCATGAACTGG TGCTGATTAATG
        AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCtaGAACTGG TGtcGATTAATG
        AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCATGAACTGG TGCTGATTAATG
        AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCATGAACTGG TGCTGATTAATG
        AGTGTAGCTTTACCGGCTACCGGG GACGTATCGGCATTCATGAACTGG TGCTGATTAATG
ex      AGTGTAGCTTTACCGGCTACCGGG  ACGTATCGGCATTCATGAACTGG TGCTGATTAATG
              TTAC G CTACCGG   ACGTtaCG CATTCATGAACTG  TGCT ATTAATG
                    GGCTA  GGG  ACGTATCGGCATTCATGAACTGG TGCTGATTAATG
                           GGGaGACGTATCGGCATTCATGAACTGGqTGCTGATTAATG
                                ACGTATCGGCATTCATGAACTGG TGCTGATTAATG
                                ACGTATCGGCATTCATGAACTGG TGCTGATTAATG
                                      CGGCATTCATGAACTGG TGCTGATTAATG


        3550        3560        3570        3580        3590        3600
1       ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG

 1      ACGACGTCCGTGCCGCGmTCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACGT
        ACGACGTCCGTGCCGCGtaCCACCGCAGTGACGGTGAGtaGGCGtTaGCAC
        ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACnTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCT
        ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
ex      ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        AC ACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACGTC GTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACGTC GTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
        ACGACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
           ACGTCCGTGCCGCGATCCACCGCAGTGACGGTGAGATGGCGATTGCACAGATTCTGG
                                                TTGCACAGATTCTGG
```

```
       3610        3620        3630        3640        3650        3660
GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA

GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
< GGGGGAGTCGAACCACCATTCGTCAGGAC
GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACT
< GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
< GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGA TAAGGTACTGGCGGGGCTCACGA
GGGGGAGTCG  CCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
GGGGG
GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
< GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
< GGGGGAGTCGAACCACCATTCGTCAGGACGnGTTGAATAAGGTACTGGCGGnGCTCACGA
GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGatCTGGCGGGGCTCACGA
< GGGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
<  GGGGAGTCGAACCACCATTCGTCAGGACGGGTTGAATAAGGTACTGGCGGGGCTCACGA
<                                ACGGGTTGA TAAGGTACTGGCGGGGCTCACGA


       3670        3680        3690        3700        3710        3720
CCTGGGAAGAAGTGATCCGCGTAACCAAAGAGGAATGATATGGCACAGTACCACTATCAG

CCTGGGAAGAAGTGATCCGCGTAACCAAAGAGGAATGATATGGCACAGTACCACTATCAG
< C
< CCTGGGAAGAAGTGATCCGCGTAACCAAAcAGGAATGATATGGgACAGTACCACTATCA
CCTGGGAAGAAGTGATCCGCGTAACCAAAGAGGAATGATATGGCACAGTACCACTATCAG
CCTGGGAAGAA
< CCTGGGAAGAAGTGATCCGCGTAACCAAAGA
< CCTcccAAcAA
CCTGGGAAGAAGTGATCCGCGTAACCAAAGAGGAATGATATGGCACAGTACCAC
< CCTGGGAAGAAGTGA
< CCTGGGAAGAAGTGATCCGCGTAACCAAA
< CCTGGGAAGAAGTGtaCCGCGTAACCAAAGAGGAATGATATGGCACAGTACCACTATCAG
                                                        ATCAG


       3730        3740        3750        3760        3770        3780
GCGCTGGATGCGCAGGGGAAAAAATGCCGTGGCACTCAGGAGGCCGACTCTGCCAGACAG

GCGCTGGATGCGCAGGGGAAAAAATGCCGTGGCACTCAGGAGGCCGACTCTGCCAGACAG
<  CGCTGGATGCGCAGGGGAAAAAATGCCGTGGCACTCAGGc
GCGCTGGATG
<  CGCTGGATGCGCAGGGGAAAAAATGCCGTGGCACTCAGGAGGCCGACTCTGCCAGACAG
GCGCTGGAT CGacGGGGAAAAAATGCCGTGn ACTCAGGAGGCCGACTCTGCCAGACAG
     TGGATGCGCAGGGGAAAAAATGCCGTGGCACTCAGGAGGCCGACTCTGCCAGACAG
```

```
       3790        3800        3810        3820        3830        3840
GCGCGTCAGCTATTGCGGGAGCGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC

GCGCGTCAGCTATTGCGGGAGCGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC
GCGCGTCAcgTATTGCGGGAcgGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC
GCGCGTCAGCTATTGCGGGAGCGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC
       CAGCTATTGCGGGAGCGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC
            ATTGCGGGA CGCGGGCTGGTGCCGCTGTCGGTTGATGAAAACCGAGGC
                                              TTGATGAAAACCGAGGC
                                                 AAAACCGAGGC
```

```
       3850        3860        3870        3880        3890        3900
GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC

GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC
GACCAGCAcAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCA
GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC
< GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC
GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC
GACCAGCAGAAATC GGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGG TCAGCACC
< GACCAGCAGAAATCCGGTTCTACAGGACTCTCCCTGCGCCGAAAAATTCGGCTCAGCACC
<                                                       ACC
```

```
       3910        3920        3930        3940        3950        3960
TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTG GCGGCGTCGATGCCGCTG

TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTG GCGGcGTCGATGCCGCTG
TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTAC
TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTG GCGGnGTCGATGCCGCTG
TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTG GCGGcGTCGATGCCG TG
TCAGATTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTG GC
TCAGATTTGGCGC ATTGACTCGCCAGTTGcgTACGCTGGTGnGCGG  TCGATGCCGCTG
TCAGATTTGGCGCTATTGACTCGCCAGTTn   ACGCTGGTGnGCGG  TCGATGCCGCTG
     TTTGGCGCTATTGACTCGCCAGTTGGCTACGCTGGTG GCGGcGTCGATGCCG TG
                                              TCGATGCCGCTG
                                              TCGATGCCGCTG
```

```
        3970        3980        3990        4000        4010        4020
GAAGAAGCGCTGGATGCGGTGGCAAAAC AGAGCGAAAAGCCAC ATCTGAGCCAGTTGATG

1       GAAGAAGCGCTGGATGCGGTGGCAAAAC AGAGCGAAAAGCCAC ATCTGAGCCAGTTGATG
<       GAAGAAGCGCTGGATGCGGTGGCAAAAC AG
        GAAGAAGC
<       GAAGAAGCGCTGGATGCGGTGGCAAAAC AGAGC
<       GAAGAAGCGCTGGATGCGGTGGCAAAAC AGAGCGAAAAGCCAC ATCTGAGCCAGTTGA
        GAAGAAGCGCTGGATGCGGTGGCAAAACgAGA CGAAAAGCCACtATCg
<       GAAGAAGCGCTGGATGCGGTGGCAAAAC AGAGCGAAAAGCCAC ATCTGAGCCAGTTGATG
<       GAAGAAGCGCTGGATGCGGTGGCAAAAC AGAGCGAAAAGCCAC ATCTGAGCCAGTTGATG
                TGGATGCG TGGCAAAAC AGAGCGAAAAGCCAC ATCTGAGCCAGTTGATG
```

```
        4030        4040        4050        4060        4070        4080
GCGGCGGTGCGCAGCAAGGTCATGGAAGGTCACTCGCTGGCCGATGCCATGAAATGCTTT

1       GCGGCGGTGCGCAGCAAGGTCATGGAAGGTCACTCGCTGGCCGATGCCATGAAATGCTTT
<     GCGGCGG
<     GCGGaGGTGCGCAGnAAGGTCATGGAAGGTCA
      GCG CGGTGCGCAGCAAGGTCATGGAAGGTCACTCGC1GGCCGATGC ATGAAATGCTTT
<                                 GGTCACTCGCTGGCCGATGCCATGAAATGCTTT
                                                    AAATGCTTT
```

```
        4090        4100        4110        4120        4130        4140
CCCGGCAGCTTTGAACGGCTGTACTGCGCGATGGTTGCCGCAGGC GAGACATCCGGTCAC

1       CCCGGCAGCTTTGAACGGCTGTACTGCGCGATGGTTGCCGCAGGC GAGACATCCGGTCAC
        CCCGGCAGCTTTGAACGGCTGTACTGCGCGATGGTTGCCGC
<       CCCGGCAGCTTTGAACGGCTGTACTGCGCGATGGTTGCCGCAGGC GAGACATCCGGTCAC
        CCCGGCAGCTTTGAACGGCTGTACTGCGCGATGGTTGCCGCAGGC GAGACATCCGGTCAC
                    CGGCTGTACTGCGCGATGGTTGCCGCnG     AGACATCCGGTCAC
                                CGATGGTTGCCGCAGGCtGAGACATCCGGTCAC
                                                    AC
```

```
      4150        4160        4170        4180        4190        4200
CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGCCAGCAGATGCGCAGC

CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGCCAGCAGATGCGCAGC
CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGCCAGCAGATGCGCAG
CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGCCAGCAGATGCGCAGC
CTTGATGCCGTGTTGAACCGGTTAGCCGACTAC CCGAGCAGCGCCAcgAGATGCGCAcg
CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGnnaGCAGATGCGCAGC
CTTGATGCCGTGTTGAACCGGTTAGCCGACTACACCGAGCAGCGCCAn AGATGCGCAnC
```

```
      4210        4220        4230        4240        4250        4260
CGCATCCAGCAGGCGATGATTTACCCGTGCGTGTTAACCGTGGTCGCGATTGCCGTGGTC

CGCATCCAGCAGGCGATGATTTACCCGTGCGTGTTAACCGTGGTCGCGATTGCCGTGGTC
CGCATCCAGCAGGCGA
CGCATCCAGCAGGCGATGATTTACCCGTGCGTGTTAACCGTGGTCGCGATTGCCGTGGTC
CGCATCCAGCAGGCGATGATTTACCC
 GCATCCAGCnnGCnATGATTTACCCGTGCGTGTTAACCGTGGTCGCGATTGCCGTGGT
                          CCCGTGCGTGTTAACCGTGGTCGCGATTGCCGTGGTC
```

```
      4270        4280        4290        4300        4310        4320
AGCATTTTGCTGTCCGTCGTGGTGCCGAAAGTTGTTGAGCAATTTATTCATATGAAGCAG

AGCATTTTGCTGTCCGTCGTGGTGCCGAAAGTTGTTGAGCAATTTATTCATATGAAGCAG
AGCATTTTGCTGTCCGTCGTGGTGCCGAAAGTTGTTGAGCAATTTATT
AGCATTTTGCTGTCCGTCGTGGTGCCGAAAGTTGTTGAGCAATTTATTCATATGAAGCAG
 GCATTTTGCTGTCCGTCGTGGTGCCGAAAGTTGTTGAGCAATTTATTCATATGAAGCAG
                          TTGTTGAGCAATTTATTCATATGAAGCAG
                                                    GCAG
```

```
        4330        4340        4350        4360        4370        4380
GCGCT GCCGCTTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA

GCGCT GCCGCTTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
nnGC   C GCTTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
GCGCTtGCCGCTTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
GCGCT GC GCTTTCAACCCGCGTATTGATGGGG TGAGCGATGCC
 CGgT GCCGCTTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
           TTTCAACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
               AACCCGCGTATTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
                   atcccTTGATGGGGATGAGCGATGCCGTGCGCACGTTTGGA
                                                 ACGTTTGGA
```

```
      4390        4400        4410        4420        4430          4440
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCCGCGTGATGCTGCG G CAG

 CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCCGCGTGATGCTGCG G CAG
CCGTGGATGCTGCTGGCATTGCTGGCGnGATTTATnG  TTC GCGTGATGCTGCG G CAG
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCCGCGTGATGCTGCG G CAG
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATcG GTTCCGCGTGATGCTGCG G CAG
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCC
CCGTGGATGCTGCTGGCATTGCTG CGGGATTTAT
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCCGCGTGtaGtcGCG G CAG
CCGTGGATGCTGCTGGCATTGCTGGCGGGATTTATGGCGTTCCGCGTGATGCTGCG G CAG
     tATGCTGCTGGCATTGCTGGCGGGATTTATG CGTTCCGCGTGATGCTGCGcGgCAG
                TTGCTGGCGGGATTTATGGCGTTCCGCGTGATGCTGCG G CAG
                 CTGGCGGGATTTATGGCGTTCCcgGTGATGCTGCG G CAG
                   TGGCGGGATTTATGGCGTTCCcgGTGATGCTGCG G CAG
```

```
      4450        4460        4470        4480        4490        4500
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA

GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA
GAAA
GAAA
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTAC GCTGATCGGCC
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACC
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGG
GAAAAGCGACGCGTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA
GAAAAGCGACnnnnCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA
GAAAAGCGACGn  CAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCCGCATA
          GTCAGTTTTCATCGACGCTTGCTGCATTTACCGCTGATCGGCn CATA
                       GCTTGCTGCATTTACCGCTGATCGGCCGCATA
                                                CGGCCGCATA
```

```
        4510        4520        4530        4540        4550        4560
GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG

GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG
GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG
GgcCGTGGCTTAAACACGGCACGTTACGCGCGGAnnnTCAGTATTCTTAACGCCAGTGCG
GgcCGTGGCTTAAACACGGCACGTTACGCGCGGACcgTCAGTATTCTTAACGCCAGTGCG
< GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTn G
< GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG
< GCGCGTGGCTTAAACACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG
<                  CACGGCACGTTACGCGCGGACGCTCAGTATTCTTAACGCCAGTGCG
<                                    ACGCTCAGTATTCTTAACGCCAGTcgG
```

```
        4570        4580        4590        4600        4610        4620
GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGCCCGC

GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGCCCGC
 TGCCGCT GTTACAGG
GTGCCGCT GTTACAGGCGATGCGCATTAG CG TGATGTCATGAGCAATGACTATGCC GC
GT
< GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGC
< GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGCCCGC
< GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGCCCGC
< GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGCCn C
< GTGCCGCT GTTACAGGCGATGCGCATTAG CGGTGATGTCATGAGCAATGACTATGCCCGC
<          TcGTTACAGGCGtaGCGC TTAGaCGGTGATGTCATGAGCAATGACTATGCCn C
                                                    GCAATGACTATGCCCGC
                                                    GCAATGACTATGCCCGC
```

```
        4630        4640        4650        4660        4670        4680
CATCGGCTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCAGCCTGCACAAAGCGCTG

CATCGGCTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCAGCCTGCACAAAGCGCTG
 ATCG CTCTCTCTGGCGACG ATGCGt
< CATCGGCTCTCTCTGGCGACGGATGCG
< CATCGGCTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCAGCCT
< ATCGGCTCTCTCTGGCGACGGATGCGGTACGACAGG CGTCAGCCT
< ATCGGCTCTCTCTGGCGACGGATGCGGTACGAGAGG CGTCAGCCTGCACAAAGCGCTG
< ATCGGCTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCAGCCTGCACAAAGCGCTG
CATCnnnTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCnn    GCACAAAGCGCTG
CATCnnnTCTCTCTGGCGACGGATGCGGTACGAGAGGGCGTCnn    GCACAAAGCGCTG
                    ATGCGGTACGAGAGG CGTCAGCCTGCACAAAGCGCTG
```

```
        4690        4700        4710        4720        4730        4740
GAACAGACGGCGCTCTTCCCCCCAATGATGCGCCATATGATCGCCAGCGGTGAACGCAGC

GAACAGACGGCGCTCTTCCCCCCAATGATGCGCCATATGATCGCCAGCGGTGAACGCAGC
GAACAGACGGCGCTCTTCCCCCCAATGA
GAACAGACGGCGCTCTTCCCCCCAATGATGCGCCATATGATCGCCAGCGGTGAACGCAGC
GAACAGACGGCGCTCTTCCCCCCAATGATGCGCCATATGATCGCCAGCGGTGAACGCAGC
GAACAGACGGCGCTCTTCCCCCCAATGATGCGCCATATGATCGCCAGCGGTGAACGCAGC
d                         AATGATGCGCCATAT ATCGCCAGCGGTGAACGCAGC
d                                                 CAGCGGTGAACGCAGC
```

```
        4750        4760        4770        4780        4790        4800
GGTGAGCTGGACAGCATGCTGGAGCGGGCGGCGGATAATCAGGATCGGGAATTCAGTTCG

GGTGAGCTGGACAGCATGCTGGAGCGGGCGGCGGATAATCAGGATCGGGAATTCAGTTCG
GGTGAGCTGGACAGCATGCTGGAGCGGGCGGCGGATAATCAGGATCGGGAA
GGTGAGCTGGACAGCATGCTGGAGCGGGCGGCGGATAATCAGGATCGGGAA
<  GGTGAGC
d  GGTGAGCTGGACAGCATGCTGGAGCGGGCGGCG ATAA CAGGATCGGGAA
d  <  GGTGAGCTGGACAGCATGCTGGAGCGGCGGCGGATAATCAGGATgcGGAA
```

```
        4810        4820        4830        4840        4850        4860
CAGATGACGCTGGCGCTGGGGCTATTTGAGCCTCTGCTGGTGGTTAGCATGGCGGCGGTG

CAGATGACGCTGGCGCTGGGGCTATTTGAGCCTCTGCTGGTGGTTAGCATGGCGGCGGTG
```